The PD photographic process

L. K. H. van Beek

Work has been going on for several years at Philips Research Laboratories in Eindhoven on a new kind of photography based on a different principle from that of conventional photography. This is the PD process. The initials PD have a double significance: they relate not only to 'Physical Development', the essential feature of the process, but also to 'Philips-Dippel', the name given to the process by those who pioneered it, after the initiator of their work, the late Dr C. J. Dippel. The scientific results obtained under Dippel and his successors — H. Jonker after 1962 and the author after 1970 — have led to various applications of particular interest to an electronics-oriented company. The process combines very high sensitivity with a resolution never before attained at this sensitivity. There are many applications for the process, which can be used for making ultra-microfilm, photomasks with details smaller than one micron, fine metal gauze, delay lines for colour television and conductor patterns of any desired design. This article discusses the physico-chemical principles of the process and describes a number of industrial applications.

Introduction

To many people photography simply means taking pictures: father taking a snapshot of the family. Photography embraces a much wider field, however, and includes other processes with features that are particularly useful in scientific and technical applications and are not found in the more everyday kind of photography. One of these is the PD process, in which the final image is formed by *physical development* (PD stands for physical development).

What is meant by physical development is illustrated in *figs. 1* and *2*, where the conventional process of black-and-white photography is compared with the PD process. The conventional film consists of a base of paper or cellulose triacetate, coated with an emulsion of finely divided grains of silver bromide in gelatin (fig. 1*a*). On exposure, silver nuclei are formed in the grains exposed to light (fig. 1*b*). During development the grains containing silver nuclei are converted to silver by means of a reducing agent (fig. 1*c*). A characteristic of the conventional process is that the silver bromide acts both as a photosensitive reagent and as

the source of silver during development. In the PD process, on the other hand, separate substances are used for these two functions. The first substance is uniformly distributed — in molecular form — in a coating, e.g. of gelatin, on the base. The second substance, which may be silver or another metal, is added in a suitable form to the developer together with the reducing agent. The name 'physical development' derives from the earlier belief that the metal from such a solution 'condensed' on the nuclei formed in the photosensitive reagent. Although we now know that the development in this case is also a chemical process, the old name has remained.

In principle therefore the photosensitive coating of a PD film contains no grains (fig. 2*a*). Where the film is exposed the photosensitive molecules are activated (fig. 2*b*). In an intermediate step, preceding the development, the exposed film is immersed for a fraction of a second in a dilute solution of an appropriate metal salt, causing metal nuclei to be formed in the sensitive layer by reaction with the activated molecules (fig. 2*c*). These nuclei are made to grow by allowing silver or some other metal to be deposited on them from the developer (fig. 2*d*).

Dr L. K. H. van Beek is with Philips Research Laboratories, Eindhoven.

The differences outlined above between the conventional and the PD processes determine the special features of the methods. If a high resolution is desired, the photographic material must meet three requirements: the photosensitive compound must be sufficiently dense, the grains formed by development must be sufficiently small, and there must be very little scattering of light in the material.

factory. These results, which are rather disappointing considering the extremely small grain diameter, are mainly due to the scattering of light from the grains during exposure. The scatter effect is intensified because the light rays have to travel a relatively large distance in the sensitive layer (3 to 6 μm). Thinner sensitive layers are not a practical possibility since they would not be able to contain enough silver bromide.
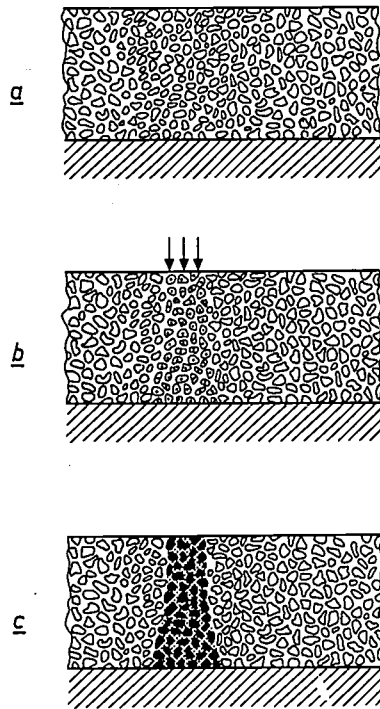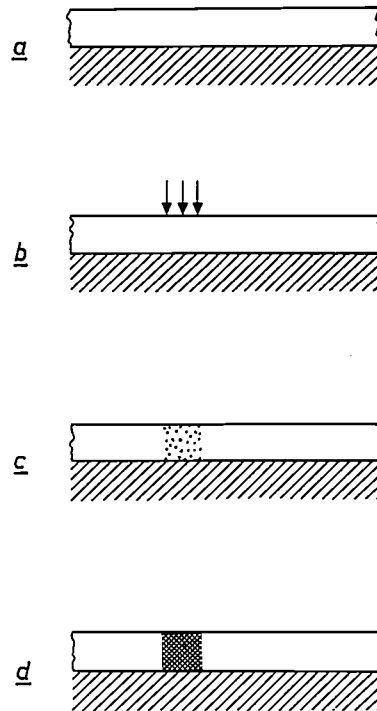
Fig. 1   Fig. 2

Fig. 1. Conventional black-and-white photography. *a*) The photosensitive layer consists of silver-bromide grains uniformly distributed in gelatin. The base (hatched) consists of paper or cellulose triacetate. *b*) On exposure silver ions are converted into silver nuclei. *c*) The grains containing the silver nuclei are completely converted into silver during development. Because of light scattering, the image spreads out in the downward direction.

Fig. 2. A typical PD system. *a*) A base material (hatched) is given a coating which is made photosensitive by a dispersion of molecules of *trans* benzenediazosulphide. (The coating can be thinner than that of a conventional AgBr film, and various materials may be used for the base.) *b*) During exposure molecules are activated: the inactive *trans* form of benzenediazosulphide is converted into the active *cis* form. *c*) In contact with a solution of silver and mercurous salts, metal nuclei are formed. *d*) During development the silver, which is contained in the physical developer in salt form, is deposited on the metal nuclei.

The AgBr grains in the layer of emulsion are always so closely packed that the first requirement is no problem, except in very thin layers. The second requirement causes more difficulties, since the usual kinds of AgBr film contain fairly coarse grains (up to 3 or 5 microns). There are however special AgBr films that have grains no greater than 0.05 μm. In practice it is found that details as small as 3 μm can be resolved fairly easily with film of this type, but smaller details present problems. Moreover the edge acuity of details that can easily be resolved is not always entirely satis-

In these respects the PD process is more promising since individual molecules and not grains are present during the exposure. In the first place this means that there are no limitations arising from grain size, and in the second place the adverse effect of scatter is virtually eliminated. What is more, the photosensitive layer can if necessary be made thinner than 1 μm, and layers as thin as this can still accommodate sufficient photosensitive reagent.

Another difference between the two processes is to be found in the developed image. On photomicrographs
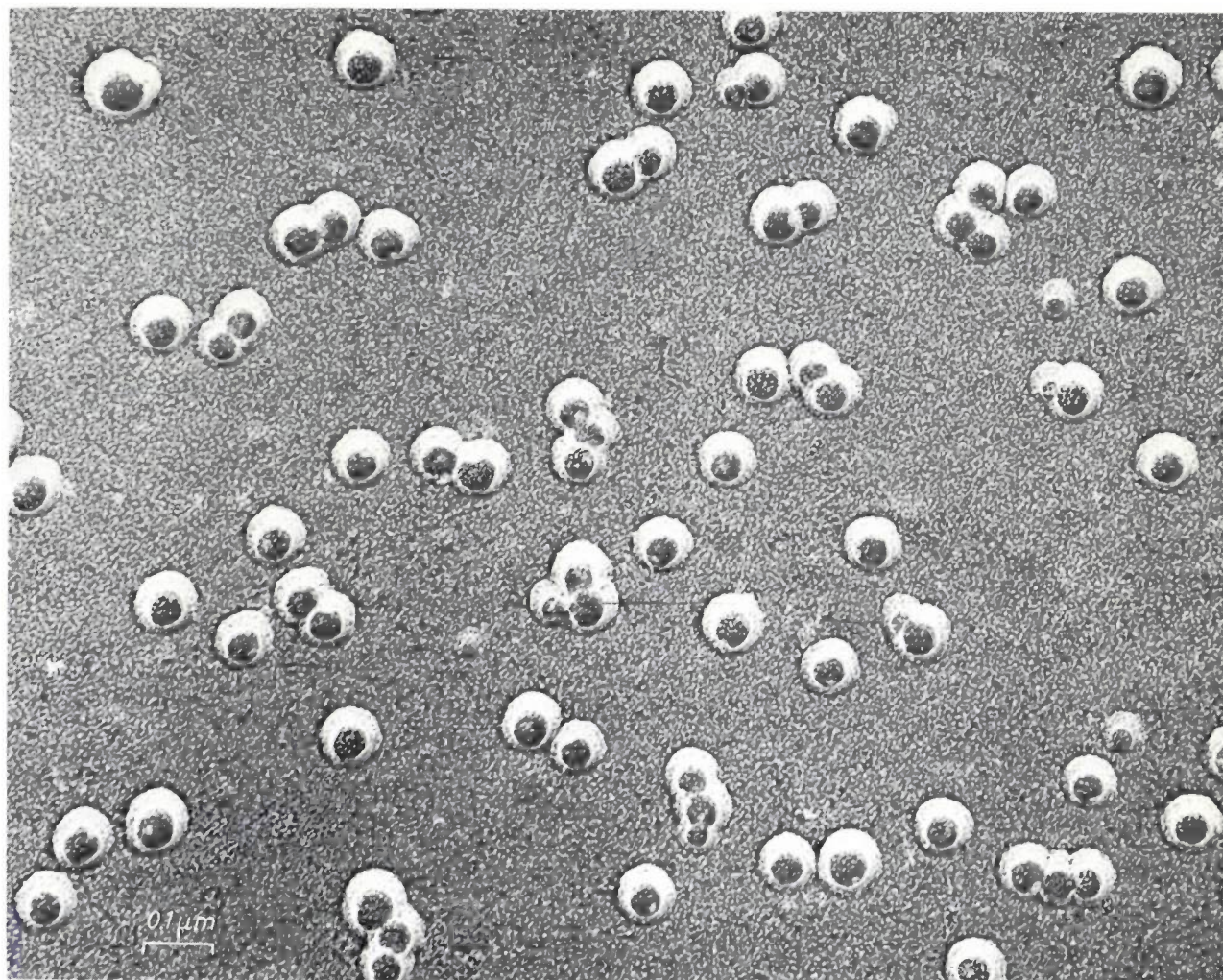
Fig. 3. In the PD process the latent-image nuclei grow into rounded particles, and not into masses of filaments as in the conventional process. This feature of the PD process also contributes to the higher resolution.

it can clearly be seen that the silver in conventional film grows in tiny filaments during the development of an exposed silver-bromide grain, resulting in a tangled mass that causes additional blurring. In the PD film, on the other hand, the silver is deposited on the nuclei in such a regular way that the grains formed are compact rounded particles (*fig. 3*).

These features make the PD film capable of resolving very fine details. By concentrating on this aspect we have succeeded in producing materials possessing such a high resolution that the edges of lines only 0.3 μm wide (this corresponds to 1500 black lines and the same number of equally wide interlinear spaces per millimetre) are still clearly enough defined. Material of this type is particularly suitable for making ultra-microfilm, and for making the photomasks used in the manufacture of integrated circuits. A special section on these two applications of the PD process is included later in this article.

Other practical applications of the PD process have stemmed from the possibility of continuing the physical development for so long that the image 'grows out of the layer' and appears in relief as a conducting metal layer on the surface of the film. The metal patterns produced photographically in this way can be reinforced with copper to form patterns of any desired thickness, for example by using the long-established electroplating processes. This combination of techniques has opened the way to an entirely new method of fabricating printed wiring.

Printed wiring is still widely produced by subtractive or by semi-additive methods, in which the deposited copper is partly etched away (*fig. 4*). A purely additive PD method for this purpose has been developed at Philips Research Laboratories; this method has fewer process steps and is more suitable for automation of the production process. Copper losses are also completely eliminated.
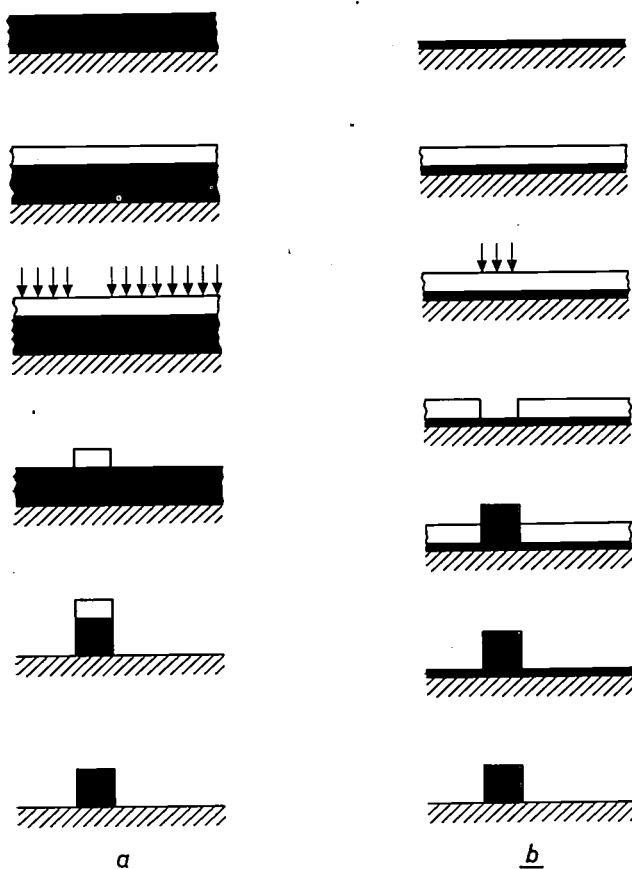
Fig. 4. Schematic illustration of two ways of making conductor patterns: *a*) by a subtractive method, *b*) by a semi-additive method (unlike these two methods, the PD process is purely additive). The base material is shown hatched, the copper black, the photoresist white. In (*a*) the exposed photoresist is removed, and the uncovered copper is then etched away. In (*b*) a very thin copper layer is applied to the base material. The unwanted parts of the copper are then screened with a lacquer, and copper is deposited on the wanted parts until a little more than the required thickness is obtained. Finally, after the lacquer has been removed, the copper is uniformly etched away until the thin, unwanted parts have disappeared.

This application of the PD process will also be discussed at greater length, but first we shall take a closer look at the method itself [1].

Since the PD materials we have developed are not panchromatic, they cannot be used for the more familiar forms of photography. The PD process uses light in the blue and near ultraviolet parts of the spectrum; the sensitivity of PD materials to this light differs very little from that of ultra-fine-grain silver-bromide materials.

The three steps of the process; reagents

Like every photographic process, the PD process begins with a reaction under the influence of light. The *direct* action of light can be made to form nuclei that react with the developer, but this has never as yet given satisfactory results. In the PD process at present used

an intermediate step is necessary. The exposed film is briefly immersed in a solution of metal salts, causing nuclei to form on the activated molecules of the photosensitive substance. The metal present in the developer is easily deposited on these nuclei. This deposition, the process of physical development, is an autocatalytic process. This means that the metal deposited, usually silver, itself acts as a catalyst for the further deposition of metal. The reactions take place in the following sequence:
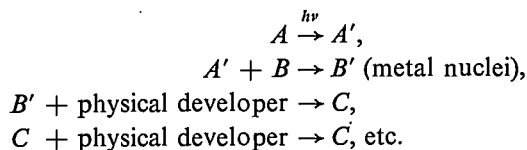
$$A \overset{h\nu}{\to} A',$$
$$A' + B \to B' \text{ (metal nuclei)},$$
$$B' + \text{physical developer} \to C,$$
$$C + \text{physical developer} \to \overset{.}{C}, \text{ etc.}$$

Table I. Three combinations of photosensitive material (A), nucleator (B) and developer metal (C).

	A	B	C
1	benzene-diazosulphide	silver and mercurous salts	silver
2	benzene-diazosulphonate	silver and mercurous salts	silver
3	titanium dioxide [2]	palladium salt	copper, nickel

Table I shows the combinations of substances A, B and C that will be discussed in some detail in this article as they are the ones that have so far been most widely used.

We shall now consider the first step of the process, and take as an example the benzenediazosulphide reaction. Under the action of light the *trans* form of this compound is converted into the *cis* form:
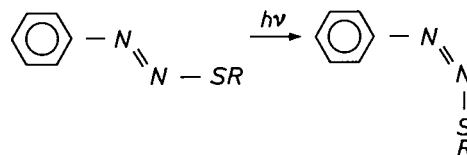
[1] The first publication on this subject was by C. J. Dippel and H. Jonker in: Reprographie (I), edited and published by Dr. O. Helwich, Darmstadt 1964, on p. 187.
A more extensive treatment of the PD process will be found in the following seven publications:
H. Jonker, C. J. Dippel, H. J. Houtman, C. J. G. F. Janssen and L. K. H. van Beek, Photogr. Sci. Engng. 13, 1, 1969;
H. Jonker, C. J. Dippel, H. J. Houtman, A. Molenaar and E. J. Spiertz, ibid. 13, 33, 1969;
H. Jonker, A. Molenaar and C. J. Dippel, ibid. 13, 38, 1969;
H. Jonker, C. J. G. F. Janssen, C. J. Dippel, Th. P. G. W. Thijssens and L. Postma, ibid. 13, 45, 1969;
H. Jonker, L. K. H. van Beek, C. J. Dippel, C. J. G. F. Janssen, A. Molenaar and E. J. Spiertz, J. photogr. Sci. 19, 96, 1971;
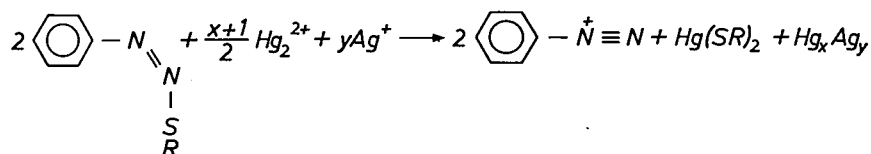H. Jonker, L. K. H. van Beek, H. J. Houtman, F. T. Klostermann and E. J. Spiertz, ibid. 19, 187, 1971 and 20, 53, 1972.
[2] Unlike the two other photosensitive compounds, titanium dioxide is not distributed in molecular form; this is discussed on page 11.

The benzene ring usually contains one or more substituents; R is frequently an alkyl group. As can be seen, the compound of the one form 'switches' to the other. (A similar reaction takes place with benzenediazosulphonate.)

In the second step of the process the *cis* compound reacts — which the *trans* form does *not* — with silver and mercurous salts, giving rise to the formation of silver amalgams, i.e. intermetallic compounds of silver and mercury:

$$2 \langle \bigcirc \rangle - N_{\backslash \backslash N \atop | \atop S \atop R} + \frac{x+1}{2} Hg_2^{2+} + yAg^+ \longrightarrow 2 \langle \bigcirc \rangle - \overset{+}{N} \equiv N + Hg(SR)_2 + Hg_xAg_y$$

where $x + y = 1$.

The third step of the process consists in each case of an oxidation reaction followed by a reduction reaction. In the oxidation reaction the reducing agent Red in the developer gives up an electron to the nucleus of silver amalgam:

$$Red \rightarrow Ox + e, \qquad (1)$$

where Ox is the oxidized form of the reducing agent Red. (If this reaction is reversible, it is referred to as a reversible redox system or redox couple; an example is Fe^{3+}/Fe^{2+}.) In the reduction reaction a metal ion M^+, in our case a silver ion, is reduced by the electron:

$$M^+ + e \rightarrow M. \qquad (2)$$

Both reactions take place at the surface of the nuclei, with the result that the nucleus grows by the addition of a silver atom, a process that is repeated many times during development.

With the two other combinations of substances the process takes place in much the same way.

We shall now examine some advantages and disadvantages of the various combinations. Benzenediazosulphonate has the advantage of being soluble in water. Benzenediazosulphide is soluble in organic solvents, but not in water. It has the important advantage of giving very little fading after exposure — the effect in which the active form partly reverts after some time to the inactive form. The photographic density obtained with benzenediazosulphonate is ten times less if after exposure nucleation and development are delayed 15 minutes; the exposure result obtained with benzenediazosulphide, on the other hand, can remain for weeks without undergoing any significant changes. With benzenediazosulphonate fading is not a disadvantage in applications where development takes place immediately after exposure.

If titanium dioxide is used as the photosensitive reagent in combination with a palladium salt as a nucleator no silver need be used at all. In this case the practical procedure differs from that with the other combinations. We shall go into this subject in more detail when describing the application based on the photo-reaction of titanium dioxide.

An effect not yet dealt with is the spontaneous nucleation that occurs in the physical developer. It is a fundamental property of the physical developer that it causes reduction of metal ions at the nuclear surface, irrespective of the manner in which the nucleus has formed. If nuclei form spontaneously in a bottle of developer, this means that complete precipitation of the metal will soon take place, making the developer useless. Not until about 1958, when the group at Philips Research Laboratories managed to eliminate this 'intrinsic' instability of the physical developer, could the PD process evolve from a curiosity into a method capable of application on a wide scale. This very important stabilization is the subject of the following section.

Stabilization of the developer

The instability of the physical developer is caused by the simultaneous presence of the reducing agent Red and the metal ions M^+ in the developer. These react with each other as indicated in reactions (1) and (2), with the transfer of electrons and the spontaneous formation of metal nuclei. The problem is therefore to keep the development rate high for the photographically formed nuclei, while at the same time avoiding spontaneous nucleation and the precipitation following it.

To solve this problem we must examine what happens during the physical development.

When a silver nucleus is brought into an environment that contains sufficient Ag^+ ions, the nucleus takes up Ag^+ ions, and with every ion it takes up it acquires a higher positive charge. As this process speeds up, the potential difference between nucleus and liquid causes the release of an increasing number of Ag^+ ions, until an equilibrium is reached in which as many Ag^+ ions are taken up as are released. In this state of equilibrium the potential difference — from now on this will just be called the 'potential' — has the value

$$E_{Ag} = E_{Ag}^0 + \frac{RT}{F} \ln [Ag^+]. \qquad (4)$$

Here E_{Ag}^0 represents the potential of the silver nuclei at

a concentration of Ag^+ ions of 1 gram ion per litre, the standard potential; R is the gas constant, T the temperature and F a charge constant.

A current of silver ions will only flow to the nucleus, and thus the nucleus will only grow, if the nucleus is given a potential that is lower ('less positive') than the equilibrium potential E_{Ag}, e.g. by the addition of a reducing agent (fig. 5).

When a silver nucleus is brought into an environment that contains a redox system, i.e. one in which reducing and oxidizing agents change reversibly one into the other, then depending on the chosen concentrations and standard potential the nucleus either takes up or gives up electrons. If it takes up electrons the nucleus acquires an increasing negative charge, until here again an equilibrium is reached. The potential at this equilibrium is

$$E_{redox} = E_{redox}^0 + \frac{RT}{F} \ln \frac{[Ox]}{[Red]} . \qquad (5)$$

A current of electrons, which is required for the growth of the nucleus, will only flow to the nucleus if it is given a potential that is higher ('less negative') than the redox potential E_{redox}, e.g. by the addition of Ag^+ ions.

When reducing agent, oxidizing agent and Ag^+ ions are simultaneously present, the potential of the nucleus assumes an intermediate value E_m, at which the current of Ag^+ ions to the nucleus is equal to the current of electrons to the nucleus, and at which the reduction of silver ions to silver at the nucleus takes place just as fast as the oxidation of the reducing agent.

The growth of the nuclei — and also the spontaneous nucleation, which is based on similar reactions — takes place when the difference ΔE between the equilibrium potentials E_{Ag} and E_{redox} is positive. The higher the positive value of ΔE the stronger the spontaneous nucleation. If ΔE were negative a current would flow in the opposite direction, and silver would then be converted into silver ions. This is an unwanted reaction, for it would make silver deposition on the photographically formed nuclei impossible, and might even cause the nuclei to dissolve. The way to prevent spontaneous nucleation is therefore to give ΔE the lowest possible positive value.

This does not necessarily imply low development rates, since the development rate depends less on ΔE than on the concentrations of the reagents, in particular the concentrations of Ag^+ and the reducing agent. A high development rate can be obtained without the unwanted consequences of strong spontaneous nucleation. Again, there is a special way to prevent this from speeding up the growth of nuclei that are still being spontaneously formed, which would be just as disastrous.

Large concentrations of Ag^+ and reducing agent give a high development rate. From equations (4) and (5) we may derive the following expression for a reversible redox couple Fe^{3+}/Fe^{2+}:

$$\Delta E = E_{Ag} - E_{Fe^{3+}/Fe^{2+}} = E_{Ag}^0 - E_{Fe^{3+}/Fe^{2+}}^0 -$$
$$- \frac{RT}{F} \ln \frac{[Fe^{3+}]}{[Ag^+][Fe^{2+}]} . \qquad (6)$$

It can be seen from this equation that ΔE can be small for large concentrations of Ag^+ and Fe^{2+} if steps are taken to ensure that the Fe^{3+} concentration is sufficiently high; in this case spontaneous nucleation will be negligible.

But however few nuclei may now form spontaneously, the rate at which they grow is so much greater that if no further measures are taken the life of the developer is still generally too short for practical purposes.

An adequate solution to the problem has been found by H. Jonker and A. Molenaar [3] who were able to use microelectrophoresis to show that the metal nuclei spontaneously formed in the developer are negatively charged. This led to the idea of screening these nuclei with positively charged surface-active molecules (soap molecules). These molecules consist of a long apolar chain with a positively charged group at one end. It
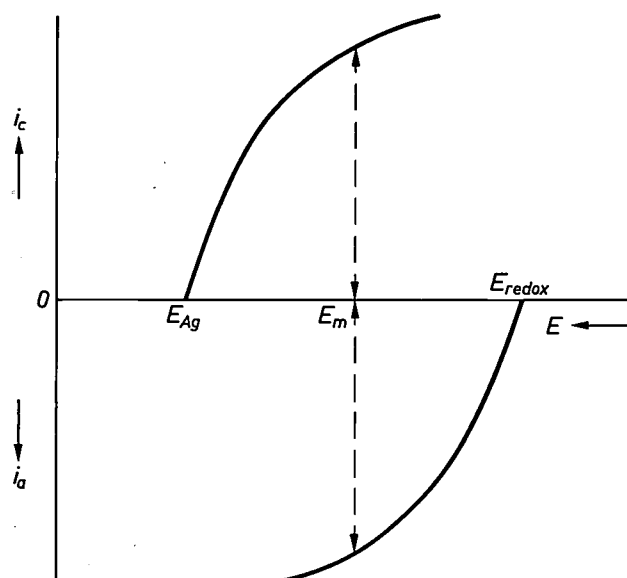

Fig. 5. The current i_c of Ag^+ ions to a nucleus and the current i_a of electrons to the nucleus as a function of the potential difference E between nucleus and solution (the 'potential'). When the silver nucleus is brought into an environment of Ag^+ ions, an equilibrium is established at E_{Ag} (zero current); when the nucleus is in equilibrium with a mixture of reducing and oxidizing molecules, the potential is equal to E_{redox}. In the simultaneous presence of Ag^+ ions and reducing and oxidizing molecules, silver is deposited on the nucleus. The potential thus acquires an intermediate value E_m at which the current of Ag^+ ions to the nucleus is equal to the current of electrons to the nucleus.

would appear that these positively charged groups gather around the negative nucleus, while the apolar chains point outwards from it (*fig. 6*). The nucleus thus surrounded will be surrounded by a second layer of soap molecules in the solvent, but this time the posi-
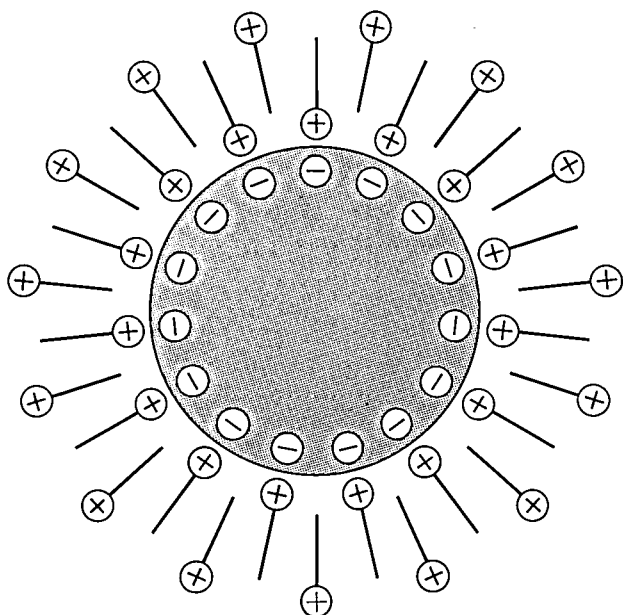

Fig. 6. The nucleus in a physical developer surrounded by a micelle consisting of stabilizer molecules. The metal cations in the developer are repelled by the micelle, which prevents the nucleus from growing further.

tively charged groups remain on the outside. This results in fairly large, positively charged particles ('micelles') on which the metal ions in the developer cannot readily settle since they also carry a positive charge.

For the physical development of a photographically formed latent image it is however essential that the stabilizing surfactant should not penetrate into the photosensitive layer, because in that case the nuclei there too would be screened. In the most commonly used materials in which the photosensitive compound is dispersed the pores are so small that the stabilizer molecules have no chance of penetrating [4].

The techniques described above have been applied in making physical developers whose life is a hundred to a thousand times longer than that of the original one, with a development rate four times greater. Whereas the life of a physical silver developer used to be about 5 to 10 minutes, the stabilized versions now have a life of more than three months, and the development rate has been increased from 0.05 to 0.23 milligrams of silver per cm^2 per minute.

In applications where a conducting pattern is to be grown on the surface it is however better not to use a stabilizer, since after some time it could have the effect

of screening the metal that comes from the layer. In a continuous technique, as used with the photoplating machine that will be discussed shortly, two-bath development is preferred. In this method the developer is produced right at the place where it is required to act; this is done by bringing together separately stored solutions of the silver salt and the reducing agent. This alternative solution of the stabilization problem is only effective in some cases, however.

Ultra-microfilm

Recent years have seen a marked growth, particularly in the United States, of the microcopying business, in which whole libraries as well as catalogues, telephone directories and other bulky printed matter of continuously changing contents are copied on microfilm (reduction ratio 20-24 : 1) or HR microfilm (reduction ratio 40-60 : 1), HR standing for high reduction. Recently UR or ultra-microfilm has also come into use (reduction ratio of at least 150 : 1) [5]. The advantage of such great reductions is most evident if regularly changing data have to be sent in large quantities by mail. The number of pages of average size that can be copied on a UR microfiche (ultrafiche) is 3000 with a reduction of 150×, compared with 600-1000 for an HR microfiche. Although the price per fiche is higher for UR microfilm than for HR microfilm, the lower postal charges compensate for this when large numbers of copies have to be sent out.

The PD process based on benzenediazosulphide is eminently suited for this application. The resolution allows even greater reduction ratios to be used — up to 500 : 1 — and these reductions can also be made in yellow light and at normal room temperature (this is not possible with the other processes). Details as small as 0.3 µm can be reproduced, making it feasible to copy such bulky works as the Encyclopaedia Brittanica — 24 volumes of 1000 pages each — on a single microfiche.

Such extreme reductions require high-quality objective lenses which are carefully corrected for blue light. Lenses of this type have been commercially available for some time.

The gamma of the photographic image (the slope of the straight-line portion of the characteristic curve of density plotted against the logarithm of the exposure) can be varied as required by varying the silver-salt/mer-

[3] See the fourth article of note [1]. (Jonker, Molenaar and Dippel).
[4] When a more porous material was used the development did not in fact proceed so readily. This difficulty was overcome by using a stabilizer molecule whose apolar chain was only 0.3 nm longer, i.e. one CH_2 group.
[5] See for example C. P. Yerkes, J. Micrographics 5, 59, 1971.

curous-salt ratio during nucleation. This makes it pos-
sible to reproduce half-tones, which may be required in
copying text with photographs.

The preparation of benzenediazosulphide film is a
very simple process. It starts with a strip of transparent
cellulose-triacetate of the type commercially available.
The strip is first dipped in an alkaline bath, which has
the effect of saponifying some of the ester groups, thus
making the surface accessible to water and alcohol. The
material is then sensitized by impregnation in an alco-
holic solution of benzenediazosulphide, and, after dry-
ing, the photosensitive film is ready for use.

The sensitivity of the film prepared in this way is of
the same order of magnitude as that of high-resolution
AgBr film.

To illustrate the capabilities of our process, *fig. 7*
shows a linear reduction by 400 : 1 of a book of 247
pages, so that the whole contents of the book occupy
no more than 8×8 mm. A single page only takes up
an area of 0.1 mm²; the letter e is 4.5 μm high. *Fig. 8*
shows the re-enlargement of part of one page, which is
clearly legible. There is no other process that can rival
this performance. This, combined with the fact that the
PD process is cheaper, makes it reasonable to predict
that it will come into very wide use for microcopying.

Photomasks

The ability to resolve details smaller than one micron
also makes the PD process of interest for the fabrica-
tion of photomasks, mainly used in the production of
integrated circuits [6]. Photomasks can briefly be de-
scribed as microphotographic negatives used for trans-
ferring patterns to the photosensitive lacquer (photo-
resist) with which the silicon slices are coated. The
photomasks are placed on the slice, and after exposure
the photoresist — depending on the type — can be
removed at the unexposed or exposed places; the silicon
is then accessible for diffusion processes. Integrated cir-
cuits are usually made by repeating this procedure a
number of times with a set of multiple photomasks.

The requirements to be met by a set of photomasks
and the method of making them with a repeating
camera have been described in a previous article in this
journal [7]. The masks in a set must be in accurate
register with each other, and the overall error must be
no more than about 0.5 μm. The material for photo-
masks must therefore have high dimensional stability:
in particular it must be insensitive to temperature varia-
tions and to changes in the relative humidity of the
atmosphere with which the masks come into contact.
To keep these variations as small as possible, the masks
are usually made in air-conditioned rooms. The only
material that possesses sufficient dimensional stability

and the required transparency is glass.

Glass can be coated with a photosensitive layer in
much the same way as described above for ultra-micro-
film. Cleaned and pretreated glass plates are coated
with a solution of cellulose triacetate by means of an
automatic spraying machine [8]. The coating is dried
and hardened by infrared radiation. The process steps
described above then follow: dipping in an alkaline
bath and impregnation in an alcoholic solution of ben-
zenediazosulphide. After rinsing, the material is cut to
the required dimensions; this must be done very accur-
ately because of the small tolerances. Glass splinters are
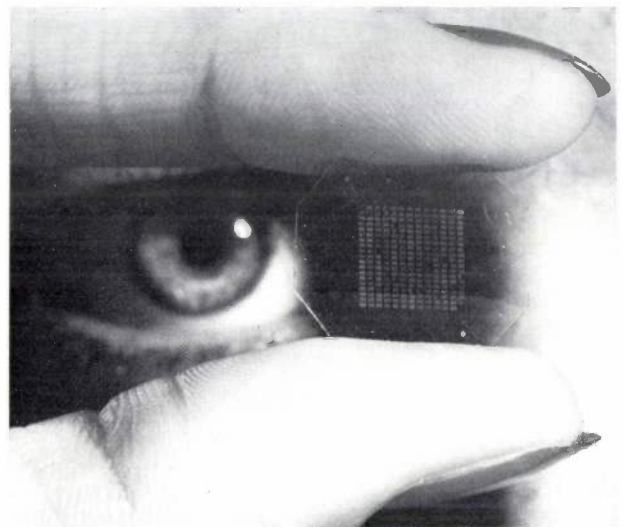
Fig. 7. A book of 247 pages (P. J. Bouman, Growth of an
enterprise, Macmillan, London 1970) reduced 400 times on PD
material.

removed ultrasonically. Since the material is not sensi-
tive to light of longer wavelengths than 500 nm, the
entire procedure can be carried out under yellow light.

In the manufacture of integrated circuits as many as
10 000 identical patterns may be recorded one after
another with a repeating camera by flash expo-
sures on a plate of 6.5×6.5 cm. This means that the
plate cannot be developed until after the recording
procedure has been completed, which may sometimes
take several hours. Here advantage is derived from the
fact that benzenediazosulphide shows no fading. Even
after a week the result of exposure in benzenediazo-
sulphide material is still present and has not changed.

It is evident that this material is suitable not only for
very fine but also for coarser masks. Reproducibility
for the various properties required is excellent, e.g. the
exposure needed for the minimum required density and
the gamma. Edge acuity and edge linearity are better
than with most other photomask materials. *Table II*
gives some data of practical interest.

It is also worth noting that positive images can be obtained by reversal processing. This extension of the potential of physical development is a fairly new development at Philips Research Laboratories. Further discussion of this process would however be out of place here.

Photoplating

Benzenediazosulphonate sensitizer

PD photoplating is our name for the method of generating metal patterns from a combination of PD

First of all the film base is given an adhesive coating (2), which for most applications must be about 4 μm thick. The adhesive is applied by passing the film over a roller which is in contact with the adhesive solution. The wet adhesive layer is then dried with warm air (3), and travels over a roller (4) in contact with a solution of benzenediazosulphonate.

This solution contains lactate additives to prevent crystallization of the diazosulphonate when it is applied to the adhesive layer. The solution is also buffered to keep the acidity constant at pH = 4 to prevent chemical decomposition, and it contains thickening additives.

To the younger members of the audience, Philips said, 'When, after having reached some conclusion, you find that someone else has a better idea, don't be headstrong; never think of opinions as a matter of prestige. It is better to turn back half way than to persist on a wrong course.'

This was a principle to which Philips had always adhered to the best of his ability: to accept advice, to listen, to take note of changing circumstances, and thus to retain the flexibility to change course in time, in short the policy of rapid adaptation so valuable to a leader. He had always been aware of the danger of professional blindness, as we have noted earlier, and he felt that it was to his wife, whose strong common sense had proved so valuable, that he owed his preservation from it. His audience found his grateful references to his wife profoundly moving.

Fig. 8. Re-enlargement of part of the ultra-microfilm of fig. 7.

photography and electroless plating or electroplating. This method was used in industrial applications at Philips as long ago as 1965. In that year a machine was put into use [9] for making wiring on a continuous strip of flexible material passed over rollers (*fig. 9*). Printed wiring boards can also be made on the same principle.

In this machine the photosensitive compound is benzenediazosulphonate, and a solution of silver nitrate and mercurous nitrate is used for forming the metal nuclei. As already indicated, a developer that contains no stabilizer is employed.

The flexible base material is usually a polyester or polyimide film. A 30 cm wide strip of it is passed over rollers as illustrated schematically in *fig. 10*. The film is transported at 5 cm per second, from a feed roller *1*.

Table II. Practical data of photographic material based on benzenediazosulphide.

Threshold exposure (405 nm)	0.25 mJ/cm²
Exposure for density $D = 2$	0.8 mJ/cm²
Exposure for electron radiation (< 20 keV)	10 μC/cm²
Gamma for photomask material (up to $D = 4$)	6
Lowest possible gamma (to $D = 2$)	0.8
Minimum line width obtainable photographically	0.3 μm

After the wet film has been dried with warm air (5), a thin glassy layer is obtained on the adhesive. With this method of application the photosensitive compound is close to the surface, so that the metal is able to build up quickly on the film.

The exposure is made through a transparent plastic drum 6, in which the light sources are fluorescent lamps mounted along the axis of the drum. A photographic negative on which the pattern is to be formed is fixed around the outside of the drum. The drum and the negative rotate together with the film, and pressure rollers ensure good contact.

[6] See for example A. Schmitz, Philips tech. Rev. 27, 192, 1966.
[7] See F. T. Klostermann, Philips tech. Rev. 30, 57, 1969.
[8] The choice of cellulose triacetate, an ordinary commercial product, represents a considerable advance over the use of gelatin in the AgBr method. The preparation of gelatin for the emulsion coating is extremely laborious, and requires a great deal of skill, understanding and experience. Even the gelatin has to be of the correct type, which can only be obtained from the hides of cattle from particular mountain pastures.
[9] This machine was developed by H. C. N. van der Sanden and H. J. Veenendaal of the Chemical Laboratory of the Philips Radio, Television and Record-playing Equipment Division, Eindhoven.
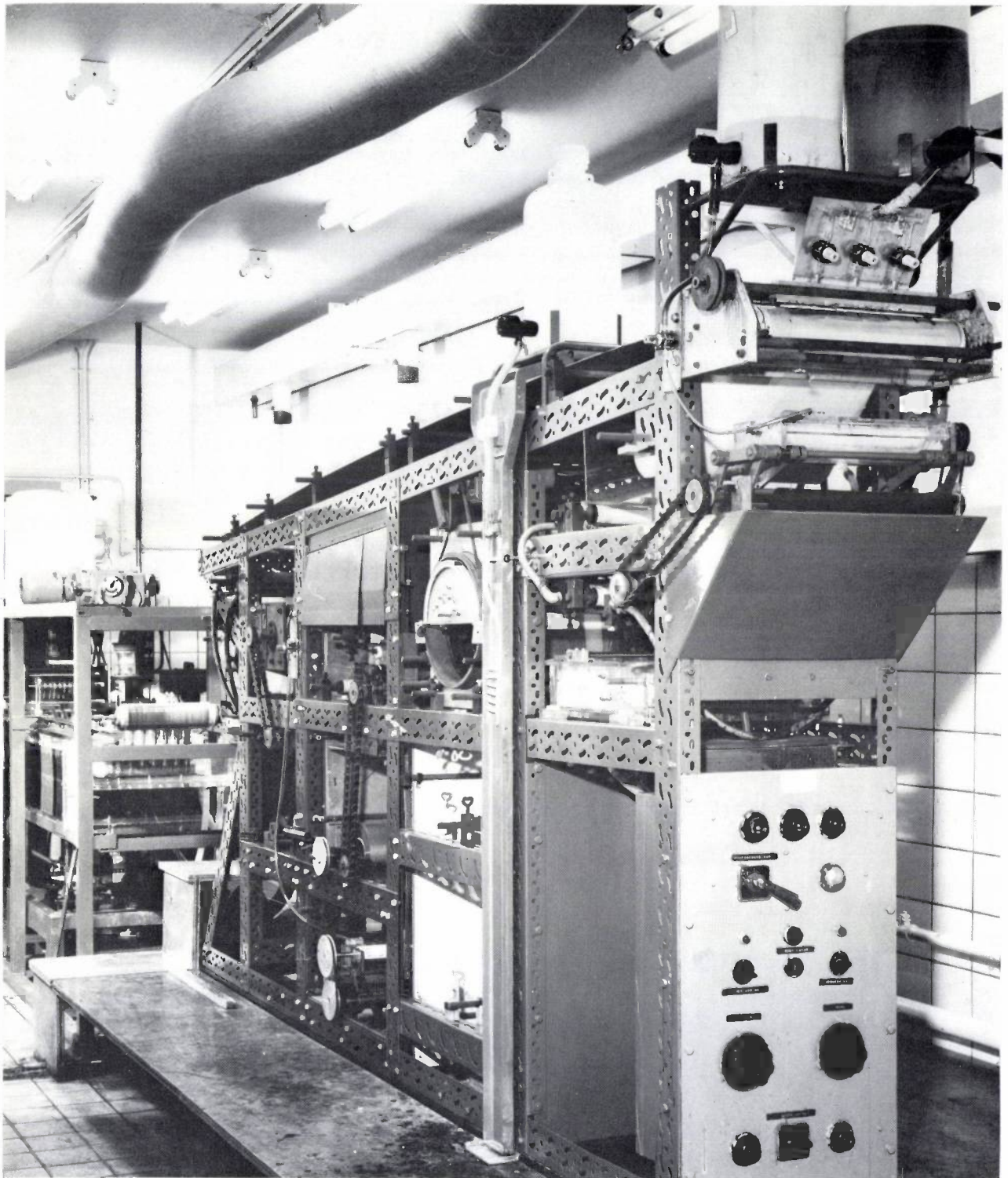
Fig. 9. A PD photoplating machine based on the use of benzenediazosulphonate as photo-sensitive reagent. The photoplating is a continuous process on a 30 cm wide strip of flexible plastic, e.g. polyester or polyimide film. The operation of the machine is illustrated in fig. 10.

The exposed film is then passed to a nucleation bath 7, a rinsing bath 8, then to a drier 9, and the actual physical development takes place on the upper track 12 (*fig. 11*). As already mentioned, this is preceded by an operation in which the silver-salt solution and the reducing agent are successively supplied from separate baths 10 and 11. During development a silver layer about 0.1 μm thick is formed. After further rinsing and drying the film is rewound (14). For many applications the conductor pattern is reinforced with the aid

of copper or nickel. This is done in an electroplating machine developed for this application (it can be seen in the background in fig. 9).

During the various rinsing operations, metal ions from the nucleation bath and from the developer are carried along with the rinsing water. Treatment with ion-exchange resins to remove these metal ions completely has been available for some time; the process does not therefore contribute to water pollution.

ductor patterns for telecommunication and space applications, fields in which reliability is very important. Silver, when subjected to strong electric fields, has a tendency to migrate, and in certain conditions this can cause short-circuiting. Although this seldom happens, the risk is enough to make the use of silver inadvisable for conductors for such applications.

It is possible to make conducting patterns without using silver if titanium dioxide is used as the sensitizer.
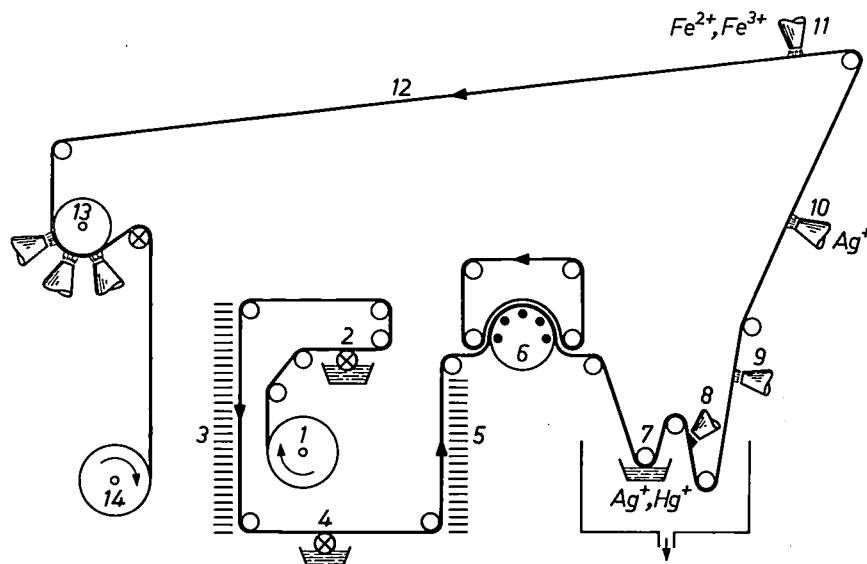

Fig. 10. Operation of the PD photoplating machine in fig. 9. *1* film supply. *2* bath containing adhesive and roller. *3* drying. *4* bath containing diazosulphonate solution and roller. *5* drying. *6* exposure drum of transparent plastic around which a photographic negative is fixed. The black dots are cross-sections of the fluorescent lamps used as light sources. *7* bath containing silver and mercurous salts and roller. *8* rinsing. *9* drying. *10* spray installation for applying silver-nitrate solution. *11* bath from which the reducing agent is applied to the strip. *12* upper track along which the development takes place. *13* rinsing and drying. *14* roller on which the finished product is wound. If necessary further plating can be carried out in an electroplating machine (this can be seen in the background in fig. 9).

Finally, there is another version of the same technique that uses the same machine. If the silver patterns are applied to a poorly binding thin coating of adhesive, the metal pattern can be stripped from the adhesive after electroplating with nickel or copper. To minimize possible damage, the adhesive can be treated for this purpose with an organic solvent. Fine wire gauze can be made in this way, e.g. for use as grids in indicator tubes. *Fig. 12* gives more information about this and other applications.

Titanium-dioxide sensitizer

In the two modifications of the PD process described above, in which the sensitizers are benzenediazosulphide and benzenediazosulphonate, silver is used for forming the image. The application may however require that no silver should be used. This is so with con-

Titanium dioxide is a semiconductor with an energy gap of 3.1 eV between valence and conduction bands. On exposure to light from the near-ultraviolet part of the spectrum, electrons can be raised from the valence band to the conduction band, thus producing an active form of titanium dioxide. This can react with palladium salts to produce palladium nuclei. These nuclei catalyse the deposition of copper and nickel in an autocatalytic physical-development process.

The photosensitivity of titanium dioxide is a property of the material in the solid state. In this case it is therefore necessary, as with AgBr, to use the sensitizer in the form of *grains*, and not in molecular forms. The resolution, compared with the two other modifications of the PD process, is therefore unavoidably lower. A special technique has made it possible to limit this reduction in resolution. The titanium dioxide grains

are applied to an adhesive layer which is mechanically or chemically roughened in such a way that the grains partly stand out from the surface. Only the parts of the grains striking out from the adhesive coating are then

with a highly diluted solution of $PdCl_2$. During exposure palladium nuclei form at the places affected by the light, and copper or nickel can be deposited on these nuclei in electroless plating baths. A machine based on
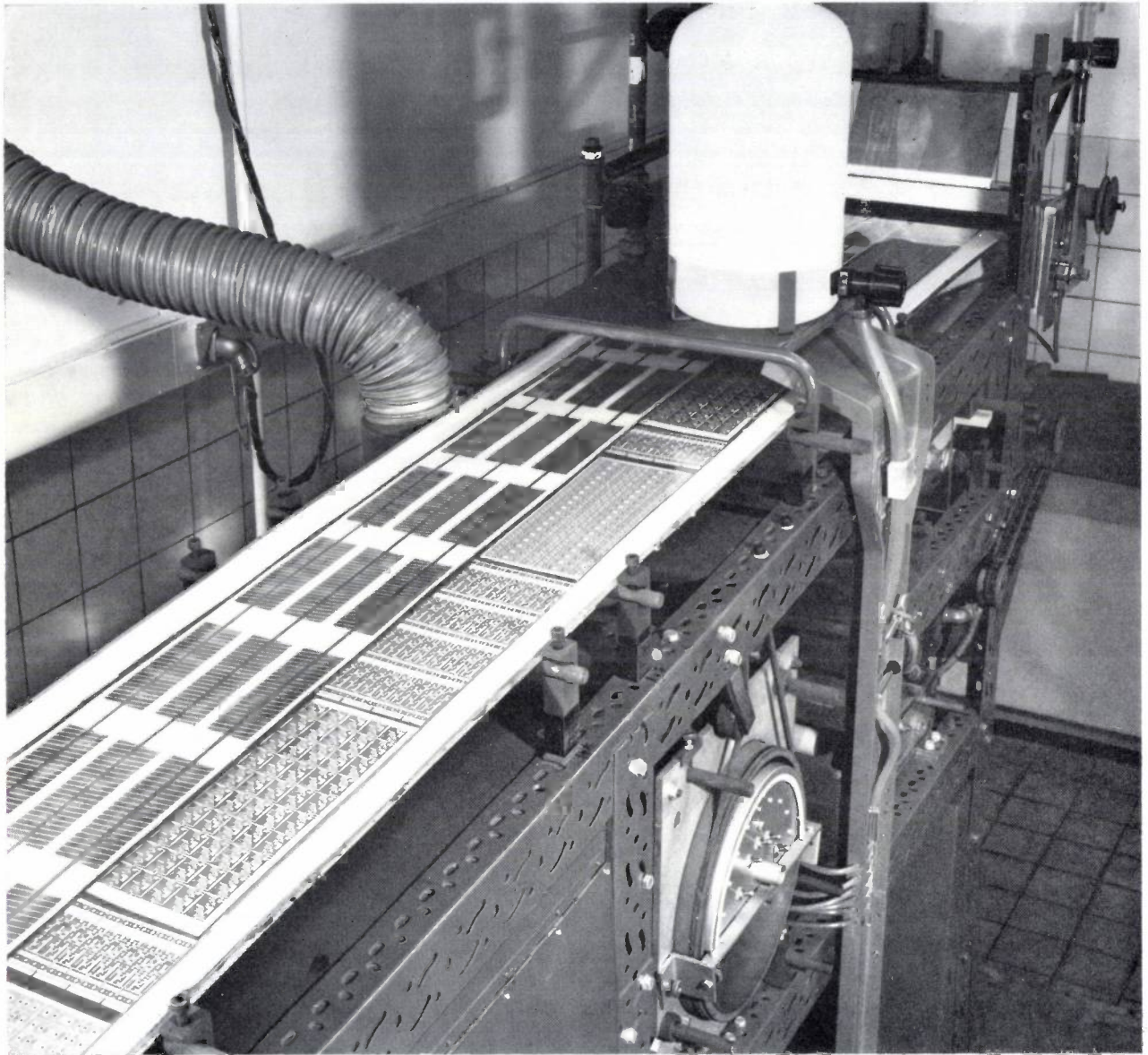

Fig. 11. The upper track of the machine in fig. 9 on which the development takes place.

activated on exposure. In the introduction it was stated that a thicker photosensitive layer implies a smaller resolution, because of light scattering. Since the process described here takes place entirely at the surface — in the same way as in the benzenediazosulphonate process — the effect of scattering is much smaller than in the AgBr film.

The film, consisting of base, adhesive coating and projecting titanium dioxide grains, is uniformly wetted

this principle, which closely resembles the machine already described, was recently put into operation at Philips.

The PD process will no doubt suggest many other applications besides those discussed in this article, which has only considered applications that are actually feasible and of interest to an electronics company. The photosensitive compounds — and there are many more

of these than mentioned here — can be combined with other substrates. This has been demonstrated with oxidized aluminium, transparent sheets of plastic, paper, cotton and wood, and other materials; in short, the PD process can be used to make photographic or metallic patterns on widely different materials. In addition *coloured* patterns can be produced by physical development. The PD process is a subject with many attractions, and indeed in our work it was often difficult to decide which interesting thing to do next.

The continuing trend towards further miniaturization of components calls for ever smaller details. The steadily increasing flood of information makes the storage of documents in greatly miniaturized form inevitable. In view of the economic necessity of automating manufacturing processes, electroless plating is an attractive alternative to vacuum-evaporation techniques.

The PD process offers new solutions to these problems, and will undoubtedly make a useful contribution to future technological progress.
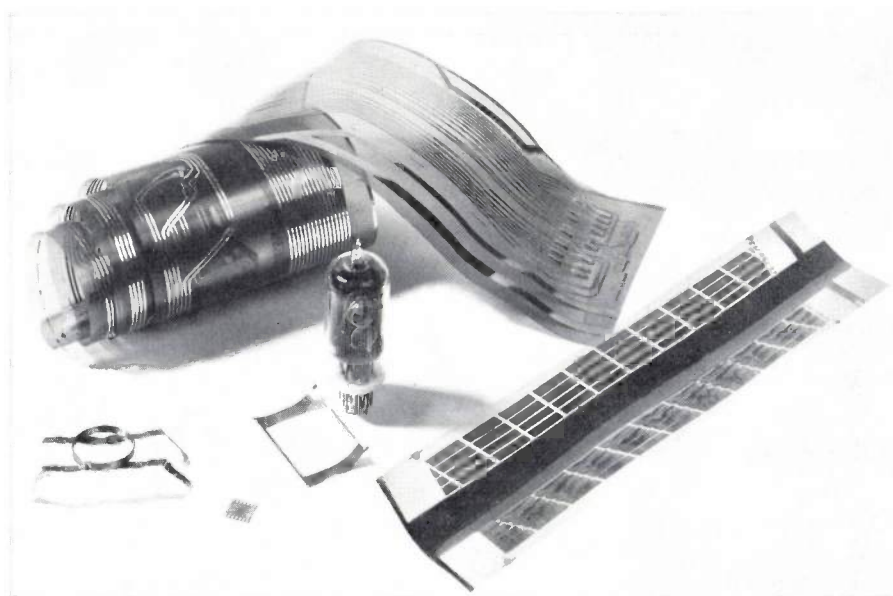
Fig. 12. Some products made by the PD photoplating process. *Right*: part of a colour-television delay line. *Left, foreground*: a microphone diaphragm. *Background*: strip with conductor pattern that can fulfil the same function as a 'tree' of connecting wires (cable-form) as used in telephone exchanges. *Centre*: an indicator tube whose anode (lying in front of the tube) is in the form of a thin mesh made by the PD process. The tiny square at the centre of the foreground is an integrated circuit on film with conductor pattern made by the PD process.

Summary. The PD photographic process described here consists of three steps: the formation of activated sites (usually activated molecules) during exposure, the formation of metal nuclei, and the further deposition of metal on these nuclei during development. The PD process differs from the conventional ones in that the metal to be deposited is contained in the developer. Since the photosensitive reagent in the PD process can be distributed in molecular form, the resolution obtainable is very high. To stabilize the metal in the developer, the difference ΔE in redox potential between the reducing redox couple and the metal is given the smallest possible positive value. At the same time, if the application allows, the (negatively charged) nuclei are surrounded by cationic surfactants.

Photosensitive reagents that can be used include benzenediazosulphide, benzenediazosulphonate and titanium dioxide. Metal nuclei are formed on exposed molecules of the first two reagents with a solution of silver and mercurous salts. Silver is used in the development. With the titanium dioxide reagent a palladium salt is used for the formation of metal nuclei, and copper and nickel are used for further development.

PD photoplating methods are also discussed in which metal patterns are built up on the photosensitive layer. In this way conductive patterns can be produced on a flexible or rigid substrate. The PD process can be used for making ultra-microfilm, photomasks, very fine metal gauze, printed wiring patterns, delay lines, etc.

Miniature pressure transducers with a silicon diaphragm

A. C. M. Gieles and G. H. J. Somers

Readers of this journal will probably now be familiar with the many advantages of the planar technology in modern electronics. It may however come as something of a surprise to learn of an application of this technology to a purely mechanical device. This is a miniature pressure transducer, whose submillimetre dimensions enable measurements to be made in situations previously considered to be completely inaccessible.

## Silicon, strain gauges and pressure meters

A pressure measurement amounts to a determination of a mechanical load. When a structure is subjected to a mechanical load, a strain is produced. It therefore follows that a pressure can be detected with a strain gauge [1] mechanically coupled to an elastic wall (such as a diaphragm) on which the pressure to be measured is exerted. The deformation of the strain gauge caused by the pressure affects the electrical resistance of the gauge. If the relation between deformation and pressure is known the unknown pressure can be determined by measuring the change in resistance.

Strain gauges used to be made — and to some extent still are — of alloy metals of high resistivity. Later it was found that single-crystal silicon is a much more suitable material for this purpose because of certain of its elastic properties, which will be dealt with presently.

In this article we shall discuss pressure transducers in which the elastic part is a minute circular diaphragm, of single-crystal N-type silicon. The diaphragm has a diameter of about 1 mm and is an integral part of a substrate acting as a clamp ring (*fig. 1*). This ring can be mounted in a simple encapsulation that also connects the transducer to the space in which the pressure is to be measured. For measuring the deformation or *flexure* of the diaphragm — and hence the pressure — the diaphragm contains four small strain gauges, which together form a complete Wheatstone bridge. The unbalance of the bridge is a sensitive measure of the pressure, provided that the gauges are located to take the best advantage of the mechanical strains in the diaphragm. In their basic design, therefore, the new pressure transducers are extremely simple [2].

An advantage of the monolithic construction is that it dispenses with the difficult and expensive process of mounting and tensioning an extremely thin diaphragm

on a separate clamp ring. Another advantage is that the strain gauges are an integral part of the diaphragm. The gauges are made by the planar technology used for integrated circuits; in fact they are simply P-type zones of high conductivity diffused into the N-type diaphragm. A separate connecting layer is therefore not needed. Such a layer could lead to troublesome creep effects, which would of course impair the stability of a transducer.
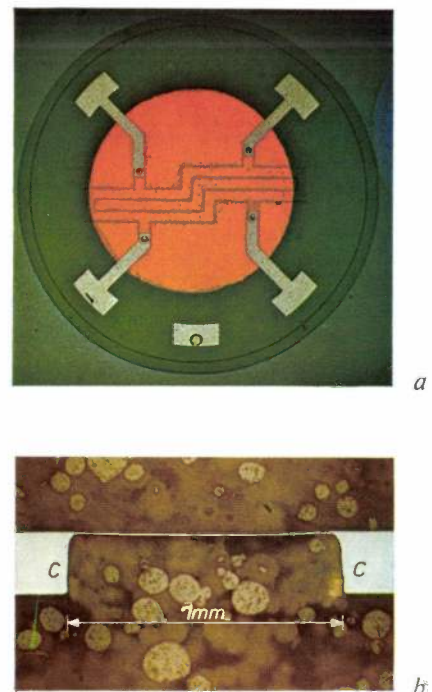
Fig. 1. The diaphragm of a miniature pressure transducer, seen from above (*a*) and in cross-section (*b*). The cross-section shows the clamp ring *C* and the circular diaphragm (typical thickness 15 μm, diameter 1 mm). Both are parts of a monolithic structure of single-crystal silicon. Diffused into the diaphragm (red circle, plan view) is an uninterrupted and electrically conducting four-active-arm pattern. The four (grey) aluminium contacts enable the diffusion pattern to be used as a Wheatstone bridge of four pressure-sensitive resistors.

Ir A. C. M. Gieles and G. H. J. Somers are with Philips Research Laboratories, Eindhoven.

The miniaturization of pressure transducers offers various advantages in measurement techniques. The first is an improved frequency characteristic, since the frequency response of a pressure transducer improves with decreasing mass and dimensions. We have made transducers that faithfully follow periodic pressure variations up to about 100 kHz. This represents a band-width about a hundred times greater than that of conventional pressure transducers. A miniaturized transducer also interferes less with the pressure to be measured. Another associated advantage is that temperature differences across the surface of the diaphragms are small.

Miniaturization has allowed many measurements to be made that were previously thought to be almost impossible, and this is of course also partly due to the useful properties of the single-crystal silicon incorporated in the transducers. Pressure measurements on scale models in wind tunnels, measurements in fast hydraulic control systems, ultrasonic measurements of certain types, and local blood-pressure measurements (for example inside a blood vessel in cardiovascular examinations) are four examples of cases where the new transducers can prove to be of great value.

Why is single-crystal silicon a more attractive material for diaphragms and for strain gauges than the alloy metals? One of the reasons is the fact that pure single-crystal material obeys Hooke's law over a wider range of strain values, the resulting stress being proportional to the strain in a given direction up to an elongation of about 1%. With the alloy metals, on the other hand, the elastic limit is exceeded at a ten times smaller strain; permanent deformation then occurs, and the reproducibility deteriorates as a result of hysteresis effects.

In structural terms, the single-crystal material is clearly superior in ruggedness and strength, and is free from hysteresis. It can withstand high temperatures, remaining perfectly elastic up to about 500 °C, and it is chemically not very reactive.

However, it is not because of any of these features that single-crystal material is used. It has been chosen because of another elastic effect: the piezoresistance effect. This is the change in the resistivity of the material caused by a mechanical stress.

From the expression $\varrho l/D$ for the resistance R of a conductor of a cross-section D, length l and resistivity ϱ, it can be shown that the change in resistance in a deformed strain gauge is composed of two terms:

$$\Delta R = \varrho\Delta(l/D) + (l/D)\Delta\varrho.$$

The first term on the right-hand side of this equation describes the change of resistance directly caused by the deformation of the conductor. The second term is the contribution of the piezoresistance effect, as the deformation is accompanied by a mechanical stress that changes the resistivity.

The alloy metals also give the piezoresistance effect, but the change it brings about in the resistance is generally considerably less than that caused by the ordinary deformation effect, i.e. by the change of length and cross-section under strain.

Since 1954 it has been known that the piezoresistance effect is very much larger than the ordinary deformation effect in single-crystal silicon [3]. In this material the same stress causes a change in resistance that is a hundred times greater than in the alloy metals, resulting in a proportionately larger output signal from a transducer using such a strain gauge. A disadvantage of silicon strain gauges is that their resistance is rather temperature-dependent.

We shall now take a closer look at the piezoresistance effect in strain gauges located in a diaphragm of single-crystal semiconducting material of cubic structure, and we shall then discuss the new pressure transducers with their characteristics and fabrication processes.

The piezoresistance effect

Let us first of all examine the physical background of the piezoresistance effect in silicon.

In a P-type single-crystal silicon chip the normal single valence band is not found, but a more complicated structure, described as a degenerate valence band. The holes are therefore of two kinds. They differ in effective mass, which has the result that the contributions to the conductivity measured in a particular direction are different. Nevertheless, the conductivity of non-deformed chips is independent of direction; this is referred to as the 'normal' conductivity. If, however, there is deformation in a particular direction, for example along a crystallographic axis, the value of the effective masses — and also the ratio between the number of 'light' and 'heavy' holes — may change considerably. Both effects are strongly dependent on the direction. The deformation is associated with a mechanical stress in the same direction. The resistivity for currents in the direction of the mechanical stress changes in that direction. The magnitude of this piezoresistance effect therefore depends closely on the direction chosen.

In the case of N-type silicon the explanation of the effect is rather different. The conduction electrons are

[1] See for example: Semiconductor and conventional strain gauges, (ed. M. Dean III and R. D. Douglas), Academic Press, New York 1962, and T. Potma, Strain gauges. Centrex, Eindhoven 1967.

[2] Further details are given in: A. C. M. Gieles, Subminiature silicon pressure transducer, 1969 IEEE Int. Solid-State Circuits Conf. Digest tech. Papers, pp. 108-109.

[3] C. S. Smith, Piezoresistance effect in germanium and silicon, Phys. Rev. 94, 42-49, 1954.

in a multi-valley conduction band, which implies that the electrons are distributed in a number of groups (e.g. six) which make different contributions to the conductivity measured in a particular direction. In this type of material the conductivity is also independent of direction as long as the material is not deformed. On deformation there may be a marked direction-dependent redistribution of the electrons among the valleys.

In order to examine the effect in a strain gauge we must know the mechanical stresses in the gauge. In our case we want to know the stress distribution in a diaphragm clamped at the edge, which is flexed by the application of pressure. The thickness of the diaphragm is much smaller than the diameter. For flexures that are small compared with the thickness the stress distribution is then very nearly circularly symmetrical in the plane of the diaphragm. The curves of the radial and tangential components of this stress distribution [4], illustrated in *fig. 2*, show for example that gauges near

the edge and near the centre are subjected to stresses of opposite sign, which can increase the output signal of a bridge circuit and thus increase the sensitivity of the pressure transducer.

The stress distributions deviate from pure circular symmetry because the elastic properties of single-crystal material are direction-dependent [5]. For example, in single-crystal silicon with a cubic structure the proportionality constant in Hooke's law is about 40% greater in the [111] direction than in the [100] direction, at least in extremely pure material. Such material is therefore much more rigid on the main diagonal than along a crystallographic axis.

The relation between the state of stress in a strain gauge that forms part of a single-crystal silicon diaphragm and the piezoresistance effect in the gauge can be described by the general expression

$$\Delta \varrho = (\pi_{\parallel} \sigma_{\parallel} + \pi_{\perp} \sigma_{\perp}) \varrho,$$

where $\Delta \varrho$ is the change in resistivity encountered by an electric current in the longitudinal axis of the strain gauge, and $\sigma_{\parallel}$ and $\sigma_{\perp}$ are the components of the stress in the diaphragm parallel to the long axis and perpendicular to it. (Any stress perpendicular to the plane of the diaphragm is neglected here.) The longitudinal and transverse piezoresistance coefficients $\pi_{\parallel}$ and $\pi_{\perp}$ can be expressed in fundamental piezoresistance coefficients (material constants) and in geometrical factors for the orientation of the normal stresses in the gauge ($\sigma_{\parallel}$ and $\sigma_{\perp}$) with respect to the crystallographic axes [6].

There are in total 36 fundamental π_{ij} coefficients ($i, j = 1, 2, \ldots, 6$); in a crystal of cubic structure only three of them are important: π_{44}, π_{11} and π_{12} [1][7]. The values of each of these three coefficients differ considerably in N-type semiconductors from the values in P-type semiconductors.

The values also depend slightly on the resistivity itself and on temperature. The coefficient π_{44} is found in situations where shear stresses occur; π_{11} and π_{12} are connected with normal stresses.

For doped silicon of resistivity no greater than a few tens of Ωcm and no less than about one thousandth of an Ωcm, the relations between the fundamental coefficients given in the first line of *Table I* are approximately correct. The expressions given in the table are found by inserting these relations in the general expressions [6] for $\pi_{\parallel}$ and $\pi_{\perp}$. The geometrical factors F_{11} and F_{12} are respectively $l_1^2 m_1^2 + l_1^2 n_1^2 + m_1^2 n_1^2$ and $l_1^2 l_2^2 + m_1^2 m_2^2 + n_1^2 n_2^2$, where l_1, m_1 and n_1 are the direction cosines of $\sigma_{\parallel}$ (i.e. of the longitudinal axis of the gauge) with respect to the three crystallographic axes, and likewise l_2, m_2, n_2, are the direction cosines of $\sigma_{\perp}$.

The behaviour of these geometrical factors, to which the anisotropy of the 'ordinary' elastic properties also relates, explains why it is possible to influence the piezoresistance effect by the choice of orientation. In our investigations the choice was limited to two orientations and to P-type silicon.

$$\sigma_r = \frac{3(p_1-p_2)R^2}{8h^2}\left[(3+\nu)\frac{r^2}{R^2}-(1+\nu)\right] \qquad \sigma_t = \frac{3(p_1-p_2)R^2}{8h^2}\left[(1+3\nu)\frac{r^2}{R^2}-(1+\nu)\right]$$
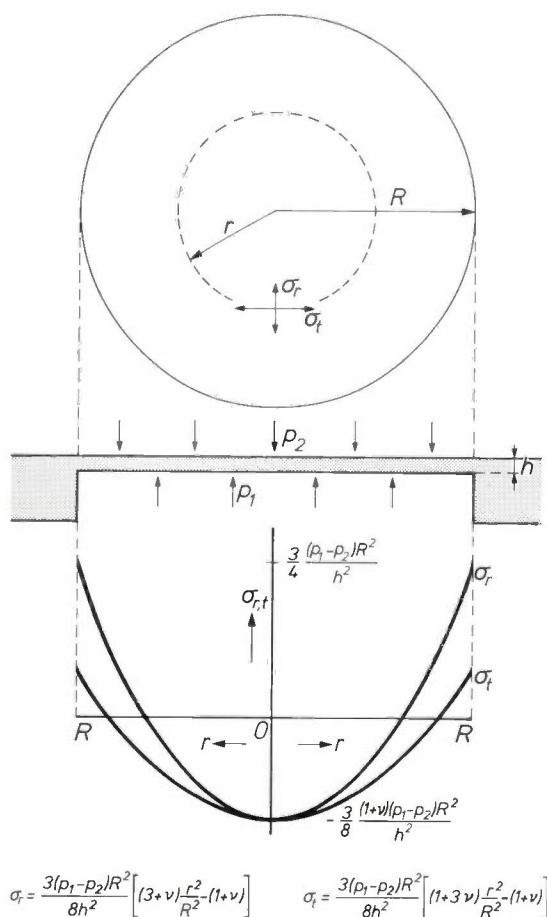
Fig. 2. The radial component σ_r and the tangential component σ_t of the stress distribution, assumed to possess circular symmetry, in a clamped circular diaphragm that flexes slightly under a uniform compressive load, plotted against the distance r between the points where these stress components occur and the centre of the diaphragm. R the radius of the diaphragm, h its thickness, p_1–p_2 the pressure difference across the diaphragm, ν Poisson's ratio.

Table I. Fundamental piezoresistance coefficients π_{44}, π_{11} and π_{12} of single-crystal silicon (resistivity between about 10^2 and 10^{-3} Ωcm) and expressions for the longitudinal and transverse coefficients $\pi_\parallel$ and $\pi_\perp$, applicable to strain gauges which are arbitrarily oriented with respect to the crystallographic axes. F_{11} and F_{12} are geometrical factors (see text).

Coefficient	Type P $\pi_{11} = \pi_{12} = 0$	N $\pi_{11} = -2\,\pi_{12}$, $\pi_{44} = 0$
$\pi_\parallel$	$2\,\pi_{44}F_{11}$	$\pi_{11}(1 - 3\,F_{11})$
$\pi_\perp$	$-\pi_{44}F_{12}$	$-\frac{1}{2}\pi_{11}(1 - 3\,F_{12})$

In the first case we choose the diaphragm plane parallel to the plane determined by a main diagonal and an edge of the cubic crystal, for example the (110) plane. For gauges in this diaphragm with a direction perpendicular to the cube edge, we then have:

$$\Delta\varrho = \tfrac{1}{2}\pi_{44}\sigma_\parallel\varrho,$$

where π_{44} is one of the fundamental coefficients. In this case the longitudinal coefficient $\pi_\parallel$ is equal to half the fundamental coefficient π_{44}, and the transverse coefficient $\pi_\perp$ is equal to zero. For gauges parallel to the edge the whole effect disappears because $\pi_\parallel$ is then zero as well as $\pi_\perp$.

In the second case, where the diaphragm plane is taken parallel to a plane determined by three side diagonals, we find on the contrary a certain isotropy. In such a plane, for example the (111) plane, the change of resistivity is given by

$$\Delta\varrho = \tfrac{1}{2}\pi_{44}(\sigma_\parallel - \tfrac{1}{3}\sigma_\perp)\varrho,$$

irrespective of the orientation of the gauge. Although in this case the piezoresistance effect is smaller than in the first case, there are definite technological advantages if the orientation of a gauge is not too critical.

There are graphs from which the fundamental coefficient π_{44} can be determined for, say, a diffusion zone in single-crystal silicon as a function of the surface doping concentration and the temperature [8]. The maximum value of this coefficient (about 140×10^{-11} m²/N at room temperature) is found for resistivities greater than about 1 Ωcm. Unfortunately the resistivity is most dependent on temperature in this region. A good compromise is a choice in the region of 0.015 Ωcm, which reduces the temperature dependence by a factor of five, while reducing π_{44} itself by only about 30%. Finally, there is the preference for gauges of *P*-type material. There are various reasons for this. One is that the fundamental piezoresistance coefficient π_{44} is about 35% greater than the coefficient π_{11} that would have to be used for *N*-type silicon. Another is that the relation between change of resistance and strain is more linear in *P*-type silicon and the temperature dependence is less.

The new pressure transducers

Location of the strain gauges

The free choice in the location of the four strain gauges on the diaphragm makes it possible to give a transducer a choice of desirable features, such as high sensitivity or temperature independence. It is of course then necessary to know the orientation of the crystallographic axes with respect to the diaphragm.

Out of the many possible locations we have only adopted the two mentioned in the previous section.

Our aim with the first location was to obtain a pressure transducer with the simplest possible bridge arrangement on the diaphragm and with the least temperature dependence. The diaphragm lies parallel to the (110) plane. The bridge consists of two gauges in the [001] direction and two in the [1$\bar{1}$0] direction, and all four are located close to the centre (*fig. 3*). The piezoresistance effect does not then arise in the first pair. The second pair, on the other hand, is subject to a reasonably strong effect. Because they are all close together the four gauges encounter very nearly identical temperature changes. A change in temperature hardly affects the balance of the bridge at all, since the four resistances then change by the same amount. The bridge does however go out of balance if the pressure changes.

Although this arrangement looks promising, it has not in practice come up to expectations, for two main reasons. The first, which is apparently related to the fairly high anisotropy of the elasticity in the (110) plane, is the presence of strongly direction-dependent mechanical stresses in the diaphragm at zero pressure. The second is that the measurements are not sufficiently reproducible, owing to variations in the junction resistances between the aluminium and the *P*-type silicon of the gauges.

These difficulties do not arise in the transducers in which the second choice of location is adopted (*fig. 4*). The diaphragms in this case are parallel to the (111) plane. The internal stresses at zero pressure are now found to be negligibly small, and the orientation of the gauges with respect to the crystallographic axes is not particularly critical. The gauges form one continuous diffusion path, and all junction resistances from alumi-

[4] S. Timoshenko and S. Woinowsky-Krieger, Theory of plates and shells, 2nd edition, McGraw-Hill, New York 1959.
[5] See the review article by H. B. Huntington, The elastic constants of crystals, Solid State Physics 7, 213-351, 1958.
[6] See W. G. Pfann and R. N. Thurston, Semiconducting stress transducers utilizing the transverse and shear piezoresistance effects, Appendix B, J. appl. Phys. 32, 2008-2019, 1961.
[7] A review of the subject has been given in: R. W. Keyes, The effects of elastic deformation on the electrical conductivity of semiconductors, Solid State Physics 11, 149-221, 1960.
[8] O. N. Tufte and E. L. Stelzer, Piezoresistive properties of silicon diffused layers, J. appl. Phys. 34, 313-318, 1963.
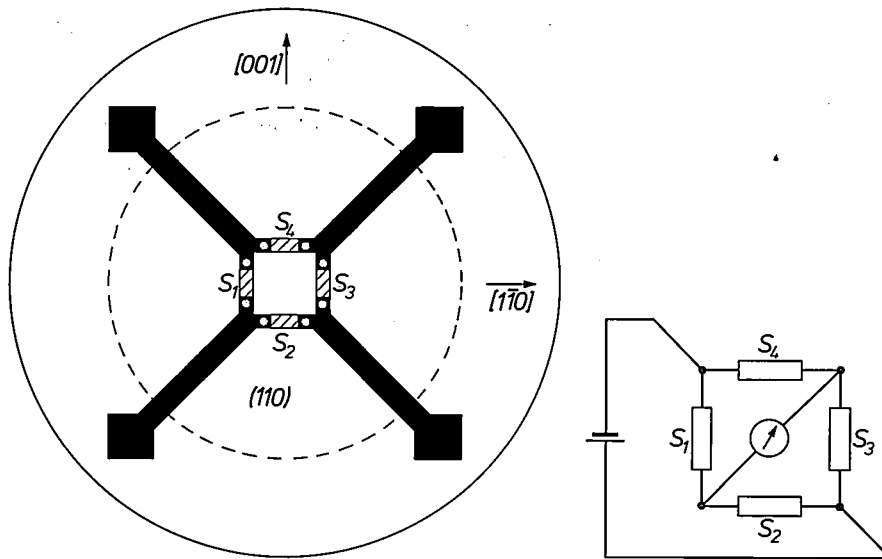
Fig. 3. Location of the strain gauges ($S_1, \ldots, _4$) in the (110) plane for minimum temperature-dependence of the pressure transducers. Gauges S_1 and S_3 react only to temperature, S_2 and S_4 to pressure and temperature. In this configuration the resistances of the junction between the aluminium connecting strips and the P-type silicon of the gauges are included in the current-carrying part of the bridge circuit, which is a disadvantage because these resistances are not constant. The area outside the dashed circle is the clamp ring, which carries the contact strips for the electrical connections.

nium to P-type silicon are kept 'outside' the current-carrying branches of the bridge circuit. This helps to prevent interference with the pressure measurements. In this configuration the four resistances are identical, and the gauges therefore have the same number of right-angle bends. The changes in resistance are of the same magnitude but of opposite sign, which increases the sensitivity of the bridge and facilitates measurements. The active parts of the gauges S_1 and S_3 are arranged radially because at the edge of the diaphragm the radial component of the stress is several times greater than the tangential component (fig. 2). Another attractive feature of the pressure transducers based on the configuration in fig. 4 is that fabrication is straightforward with few rejects.
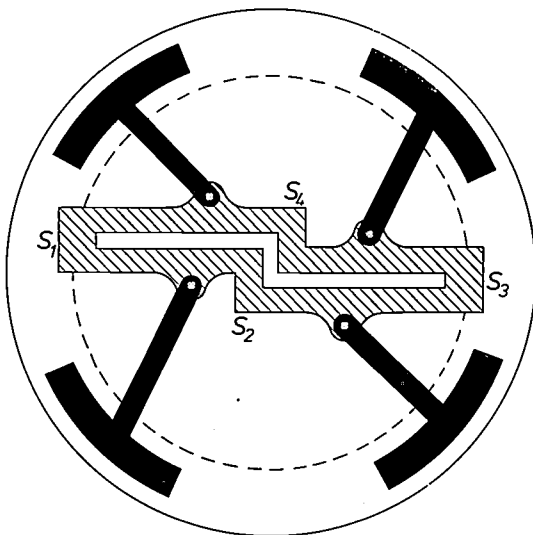

Fig. 4. Configuration of four strain gauges ($S_1, \ldots, _4$) and aluminium connections in the (111) plane. The gauges have the same resistance and the same number of right-angle bends. The changes in resistance that occur are equal in magnitude but of opposite sign. Diaphragms with this configuration are found to be more satisfactory than those in fig. 3. The dashed circle again indicates the boundary of the free part of the diaphragm.
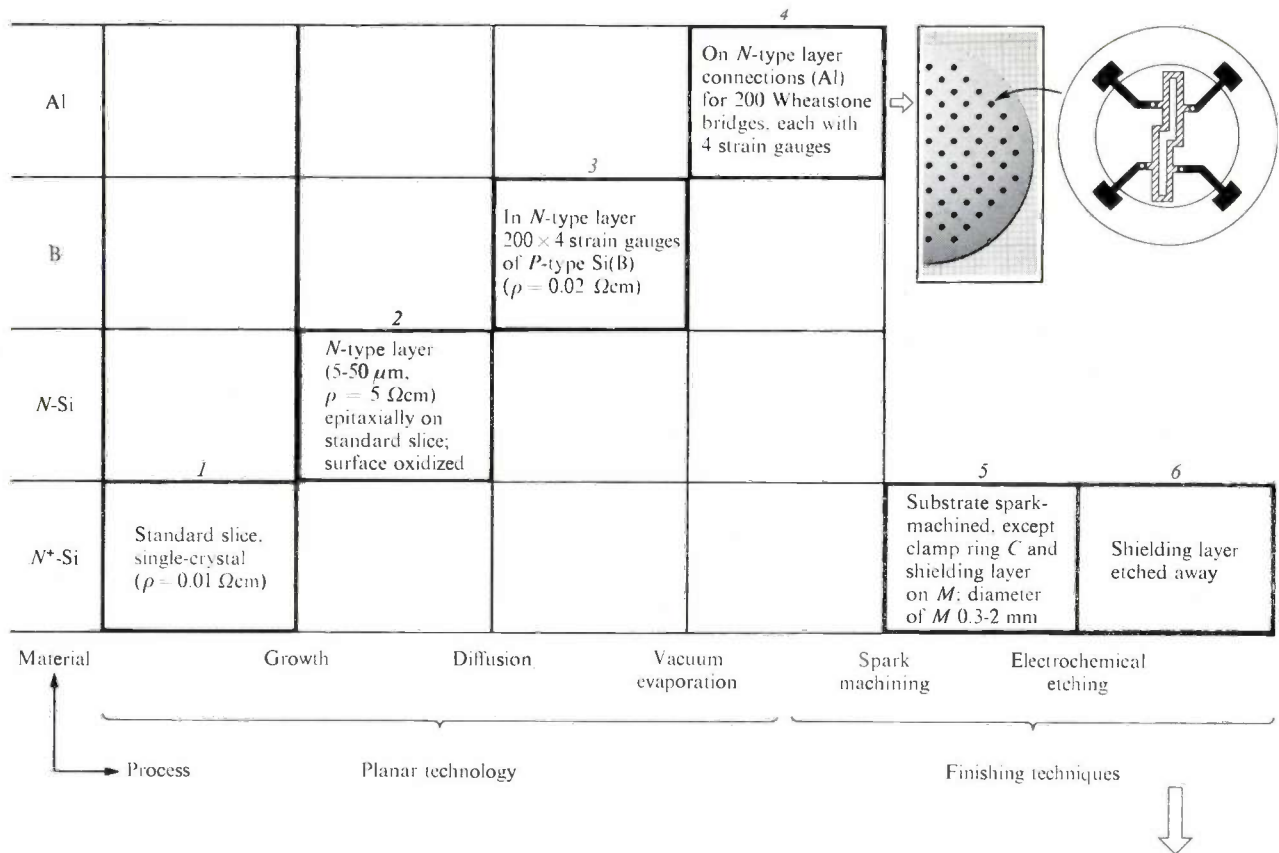
Features of the transducers

The measuring range of the pressure transducers with the configuration of strain gauges and connections illustrated in fig. 4 on the diaphragm can be varied extensively by selecting the ratio between the thickness and diameter of the diaphragms. The limits of the measuring range lie at about 1 atm and 100 atm. These values correspond to thickness/diameter ratios of 0.015 and 0.15. Within these limits the sensitivity is the same, i.e. a full-scale deflection of 20 millivolts per volt applied to the bridge. The sensitivity is satisfactorily linear with the pressure being measured, deviations remaining below 1% of maximum unbalance. If a little greater deviation from linearity is permissible, or a slightly lower sensitivity, the pressure range can be extended to 0.2 atm full scale. The linearity is worse because the flexure in the middle of the diaphragm is no longer small compared with the thickness. The diaphragm bulges out in the middle (the 'balloon effect' [4]). The transducers can safely be overloaded, and an overload of ten times full scale causes no damage. The zero shift is at the most 0.02% of full scale per degree Celsius. The adverse effect of temperature on the signal sensitivity of the circuit can be kept below about the same level by means of a compensating device, such as an NTC resistor connected in series.

The reproducibility of the measurements is excellent; the differences are no more than 0.05%, which shows that the transducers are free from hysteresis effects and creep.
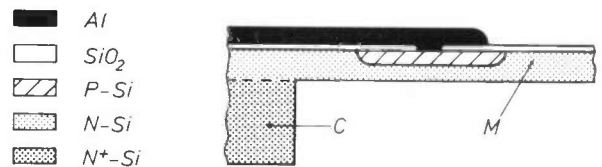
Technology

The main technological problem to be solved was to find the right combination of planar silicon technologies with two advanced finishing processes to enable the transducer with all its composite parts to be made in a single monolithic device of the required miniature

Material	Growth	Diffusion	Vacuum evaporation	Spark machining	Electrochemical etching
Al				*4* On N-type layer connections (Al) for 200 Wheatstone bridges, each with 4 strain gauges		
B			*3* In N-type layer 200 × 4 strain gauges of P-type Si(B) ($\rho = 0.02$ Ωcm)			
N-Si	*2* N-type layer (5-50 μm, $\rho = 5$ Ωcm) epitaxially on standard slice; surface oxidized					
N⁺-Si	*1* Standard slice. single-crystal ($\rho = 0.01$ Ωcm)				*5* Substrate spark-machined, except clamp ring C and shielding layer on M; diameter of M 0.3-2 mm	*6* Shielding layer etched away

Process → Material → Growth → Diffusion → Vacuum evaporation → Spark machining → Electrochemical etching

Planar technology | Finishing techniques

Table II. Survey of the technology used for the simultaneous production of 200 miniature pressure transducers. There are five steps (horizontal axis) and four materials (vertical axis). The starting material, N^+ silicon, is defined in block *1*. The blocks *2*, *3* and *4* give the results of the processing of the materials (planar technology). Blocks *5* and *6* refer to the finishing techniques, limited to the N^+ silicon substrate. Each process step takes place simultaneously for the 200 transducers, with the exception of the machining process (*4* → *5*).

Al — Al
☐ — SiO_2
▨ — P-Si
▦ — N-Si
▩ — N^+-Si

dimensions. For quantity production it was also necessary to meet the requirement of satisfactory reproducibility.

A survey of the production technology is given in *Table II*. Three conventional planar processes are used [9]. The bridge circuit can easily be kept so small that a complete signal amplifier and the bridge can be mounted on a diaphragm with a diameter of 0.5 mm.

The other techniques used, spark machining [10] and electrochemical (or anodic) etching [11], are less conventional. The first is used for removing material in the N^+-silicon substrate at the locations where the diaphragms are to be formed, and the second is used for fine finishing. The relatively high conductivity of the N^+ silicon, which has a resistivity of only 10 to 20 mΩcm, facilitates the application of spark machining. The inside diameter of the clamp ring is accurately defined by the diameter of the cylindrical electrodes. The dimensions produced so far range from about 0.3 to 2 mm, but smaller inside diameters can be produced if required. Spark machining is a relatively crude pro-

cess; the lower surface of the recesses may be pitted to depths of about 3 μm, and the centre of the recess is somewhat deeper than at the edge. Since the diaphragm (the epitaxial N layer) is usually required to be extremely thin, the spark machining is stopped when the bottom of the recess is within about 15 μm of the epitaxial layer. The remaining substrate material (the shielding layer) on the diaphragm is subsequently removed by electrochemical etching, an anodic oxidation in which the silicon-oxide layer produced is removed by a dilute HF solution. The N^+ silicon acts as the anode, and the cathode is of platinum. At an appropriate current density the N^+ silicon then

[9] The processes in question are described in: J. Goorissen and H. G. Bruijning, Doping methods for the epitaxial growth of silicon and germanium layers, Philips tech. Rev. **26**, 194-201, 1965, and in: A. Schmitz, Solid circuits, Philips tech. Rev. **27**, 192-199, 1966.

[10] See C. van Osenbruggen, High-precision spark machining, Philips tech. Rev. **30**, 195-208, 1969.

[11] A simple discussion of this technique is given in: J. A. van Nielen, M. J. J. Theunissen and J. A. Appels, MOS transistors in thin monocrystalline silicon layers, Philips tech. Rev. **31**, 271-275, 1970.

dissolves away at a rate of about 2 μm per minute, but the epitaxial layer is unaffected, because N-type silicon has a much higher resistivity and is not etched away under these conditions. The substrate layer dissolves away completely in a few minutes. At the same time the inside diameter of the clamp ring has become 20 to 30 μm larger, but this is no disadvantage as there is some rounding off at the edge between the ring and the diaphragm. The gradual transition in thickness here ensures good clamping of the diaphragm.

In this way perfectly smooth and flat diaphragms are obtained whose thickness is exactly equal to that of the epitaxial layer, and which are integral with the clamp rings. The completed transducers can easily be separated from each other, either by breaking along a pattern of scribed lines on the silicon slice, or by spark machining with a tubular electrode.

The reproducibility of the method of fabrication can be judged from the signal sensitivity of the transducers. Even with batches of several hundred transducers the variation in signal sensitivity can be kept within 10%.

The mounting of a pressure transducer is another matter. The encapsulation for such a transducer must of course be as small as possible. The encapsulation must also be designed to permit firm mechanical bonding of the four electrical leads to the bridge circuit.

Fig. 5 shows a way of meeting these requirements that is eminently suitable for applications such as blood-pressure measurements. It consists essentially of a thick-walled glass tube, with an outside diameter of about 1.5 mm and a length of 6 mm. Four thin channels are recessed in the wall for the electrical leads. The clamp ring of the transducer can be located at the front of the tube in such a way that the ends of the leads just touch the four contact strips of the transducer. These strips are located on the clamp ring (fig. 4) so that they do not load the diaphragm, and are tinned with a layer of solder. The contacts are soldered by heating the whole assembly in a non-oxidizing atmosphere, and a nickel layer along the circumference strengthens the joint and ensures a good seal. The glass tube can easily be mounted in the tip of a catheter, so that measurements can be made inside a blood vessel. What the catheter measures is of course the difference between the blood pressure and the pressure — usually atmospheric — which appears on the other side of the diaphragm via the central opening in the glass tube.

The technology described above is clearly applicable for making other types of transducer. In one such example we were able to produce an accelerometer by making a pressure transducer in which the central part of the diaphragm had a relatively large mass.
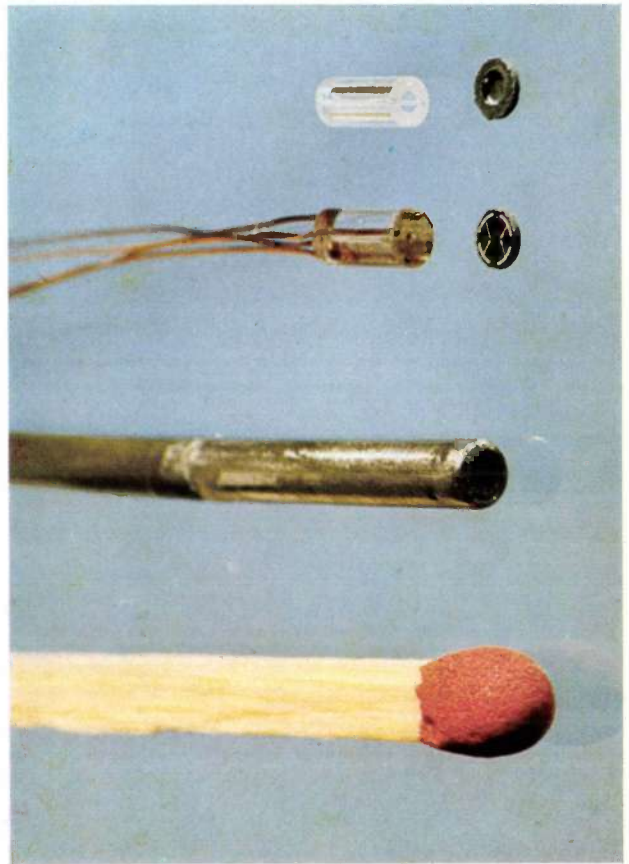

Fig. 5. The tip of a catheter for blood-pressure measurements, fitted with a miniature pressure transducer with a silicon diaphragm. The dimensions are smaller than those of a match. The various parts of the transducer are shown separately. A thick-walled glass tube supports the diaphragm with its clamp ring. The ring can be attached to the front of the tube. The wall of the tube contains recessed channels for the four electrical leads to the bridge circuit on the diaphragm.

Summary. Planar-technology processes, as used in making integrated circuits, are combined with spark machining and electrochemical etching to make miniature diaphragm pressure transducers from single-crystal silicon. The diaphragms, with diameters ranging from 0.3 to 2 mm and thicknesses from 5 to 50 μm, consist of an epitaxial layer of low-conducting N-type silicon grown on a substrate of N^+ silicon, from which the clamp ring is machined. Four pressure-sensitive strain gauges of highly conductive P-type silicon are formed by boron diffusion into the top surface of the diaphragms. The four gauges comprise a complete Wheatstone bridge.

The optimum location of the gauges to obtain a high signal sensitivity and low temperature dependence is found by applying the theory of the piezoresistance effect in P-type silicon and of stress distribution in clamped diaphragms.

The pressure ranges of the transducers vary from 1 atm to about 100 atm, in which the full-scale deflection is 20 millivolts per volt applied to the bridge. The deviation from linearity is less than 1%. An overload of ten times full scale does not affect the transducer. Bandwidths of 100 kHz and more have been obtained; the transducers are free from hysteresis and creep.

Applications include intravascular blood-pressure measurements, pressure measurements in hydraulic control systems and ultrasonic measurements.

Plasma-MIG welding

Through the years a variety of techniques have been developed for joining metals together. An important group among these techniques is that of the arc-welding processes. Recently Philips Research Laboratories have extended this group by a new method, whose essential feature is that the filler wire and its arc are surrounded by a thermally ionized gas stream. This enables the welding process to be controlled accurately, and makes possible a number of interesting applications, such as overlay welding of stainless steel on mild-steel plates at a relatively high speed and high-speed welding of thin sheet. The process may be regarded as a combination of plasma welding and gas-metal arc (MIG) welding (MIG standing for Metal Inert Gas). To make this clear, we shall first briefly describe these two processes.

In *plasma welding* the work is heated by means of an arc discharge between a non-consumable tungsten electrode and the workpiece in an atmosphere of argon or of some other gas. In plasma welding the energy density is made greater than in ordinary arc welding with a non-consumable electrode (argon-arc welding) by reducing the diameter of the arc. This is done by passing the arc through a cooled copper nozzle. The arc stays clear of the wall of the nozzle, because the gas near the wall is not ionized, owing to the low temperature, and does not therefore conduct electrically. Between the nozzle — which is part of the welding torch — and the workpiece there is very little divergence of the arc. Because the arc is thinner at the same current, the temperature of the plasma is higher [1] than it would otherwise be.

In plasma welding it is sometimes necessary to add filler metal to the weld pool. Since the tungsten electrode, the copper nozzle and the arc are all coaxial it is usual to introduce the filler wire into the arc sideways.

Filler wire introduced in this way is not very intensively heated, and the deposition rate is therefore low. The feed rate of the wire also has to be very accurately controlled. It would be better if the wire travelled a longer distance through the plasma.

MIG welding is also carried out with an arc discharge in a shielding-gas atmosphere (argon or CO_2). In this case, however, a *consumable* electrode of filler wire is used. The wire is fed continuously from a reel through a guide tube, which conducts the welding current to the wire.

Like plasma welding, MIG welding has its limitations. In MIG welding at *low* current the limit is found where the arc becomes unstable and the metal is transferred to the work in very large molten drops from the filler wire. At very *high* currents the tip of the filler wire and the arc begin to rotate rapidly, and molten drops are spattered from the arc to land well away from the heated part of the workpiece [2].

Experiments have shown that combining these two welding methods gives an entirely new situation. To indicate the close relationship we have called the new method *plasma-MIG welding* [3]. In this method there is no lower limit to the current through the filler wire below which the welding process becomes unstable, while the upper current limit is much higher than in MIG welding.
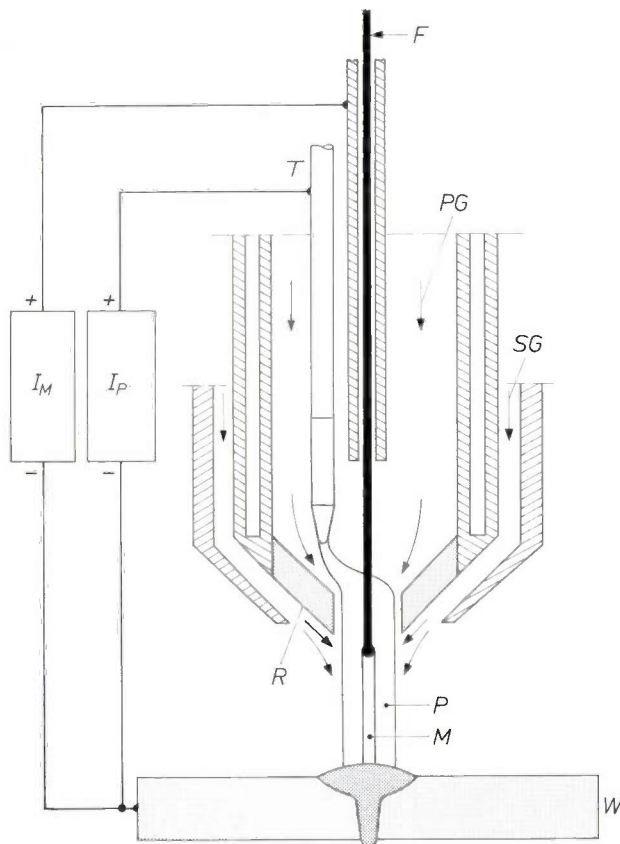
[1] H. Maecker, Z. Physik **141**, 198, 1955.
[2] To some extent the behaviour of the wire tip is comparable with that of a running hose clamped at a distance from the nozzle; when the flow of water exceeds a certain strength, the reaction causes the dangling end to start twisting. See for example the article by A. Lesnewich, Welding J. **37**, Res. Suppl., 418-s, 1958.
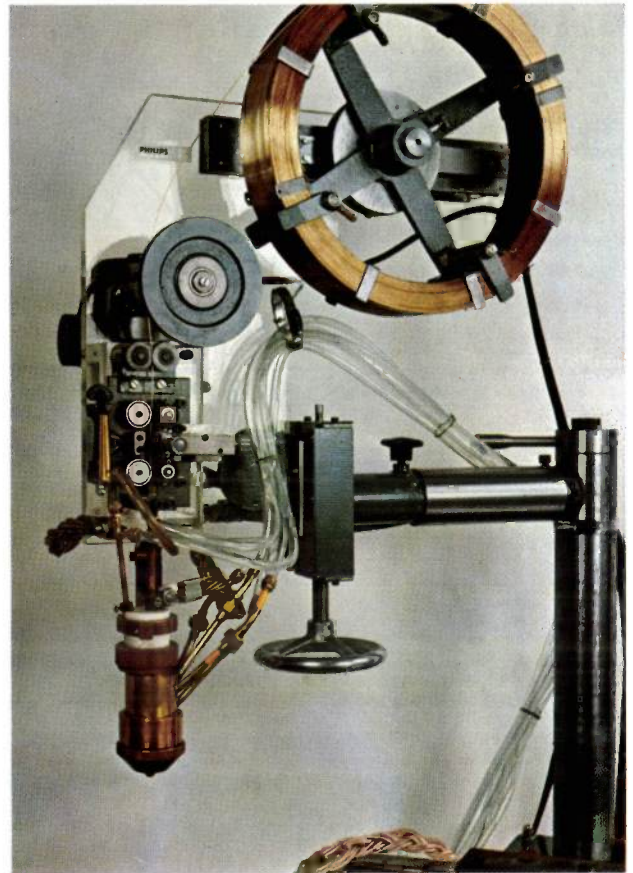[3] W. G. Essers, A. C. H. J. Liefkens and G. W. Tichelaar, Proc. Conf. on Advances in Welding Processes, 1970, p. 216, publ. The Welding Institute, Cambridge 1971.

The methods could be combined by offsetting the tungsten electrode that generates the plasma arc from the axis of the copper nozzle (see *fig. 1*). Although the arc is then slightly curved between the electrode T and the nozzle R, this presents no difficulties. The filler wire F can now be introduced into the plasma axially, and the length of wire in the plasma arc can now be much greater than with a side feed. In this way the lower part of the wire F and the arc M, between the wire tip and the workpiece W, are surrounded by the plasma P, thus ensuring efficient heating of the wire and consequently a high deposition rate. The deposition rate is of course largely determined by the current in the filler wire. The presence of the plasma also gives better control of metal and heat transfer to the workpiece than in conventional MIG welding. *Fig. 2* shows the experimental arrangement in our laboratory.

Interesting results can be obtained with plasma-MIG welding, especially when the electrodes are connected to the positive pole of the current source and the work-

Fig. 2. Experimental welding torch for plasma-MIG welding. The reel carrying the supply of filler wire can be seen at the top. The wire is automatically fed at the desired speed into the welding torch, which also contains the tungsten electrode.

Fig. 1. Diagram of the plasma-MIG welding process. An arc M is maintained between the filler wire F and the workpiece W. This arc is located at the axis of a plasma arc P between a non-consumable tungsten electrode T and the workpiece. The plasma arc is constricted by a copper nozzle R located at the end of a water-cooled double-walled housing. The plasma gas PG is supplied through this housing. SG is a stream of shielding gas around the plasma arc. The arc M and the plasma arc P are each fed from a separate d.c. source (I_M and I_P).

piece to the negative pole. Just as in ordinary MIG welding, the current in the welding wire in plasma-MIG welding has a critical value above which the molten wire tip and the arc are made to rotate by the action of electromagnetic forces and mechanical reaction forces. The process of metal transfer from welding wire to workpiece when the arc is stationary is quite different from the process when it is rotating, but in plasma-MIG welding both processes are completely controlled. We shall illustrate this with a few examples which show that the two situations lead to entirely different welding results.

Fig. 3 shows a photograph of a plasma arc with a thin stationary arc in the centre; this situation holds as long as the current in the wire remains below the critical limit. The welding process can be very effectively demonstrated by taking high-speed films and showing them in slow motion. The title photograph shows a frame from such a film (2500 frames per second) of the same arc as in fig. 3. The molten drops can be seen to be moving vertically in a straight line to the workpiece, about 500 drops per second with an
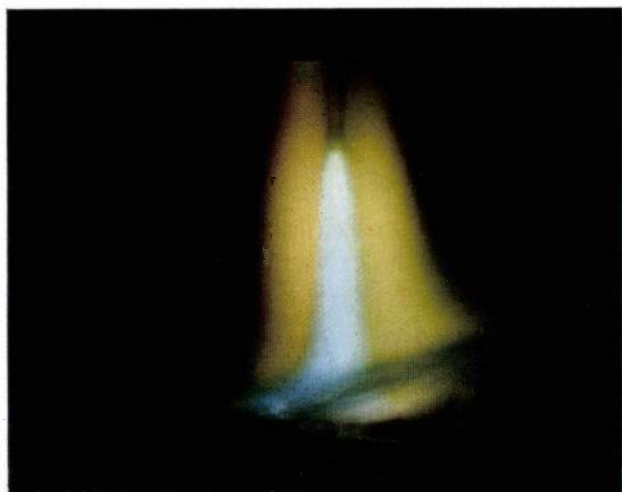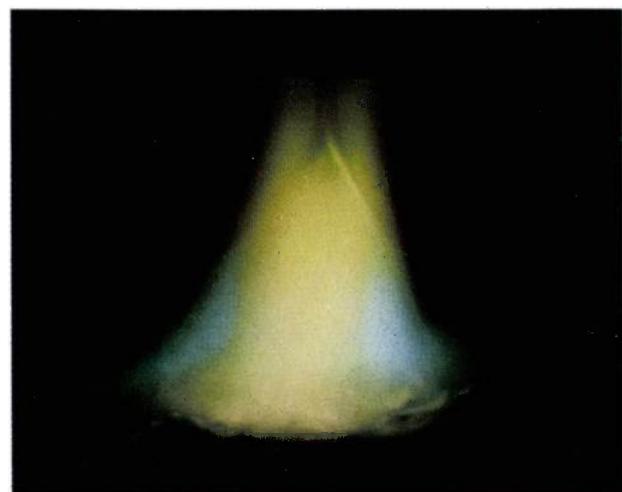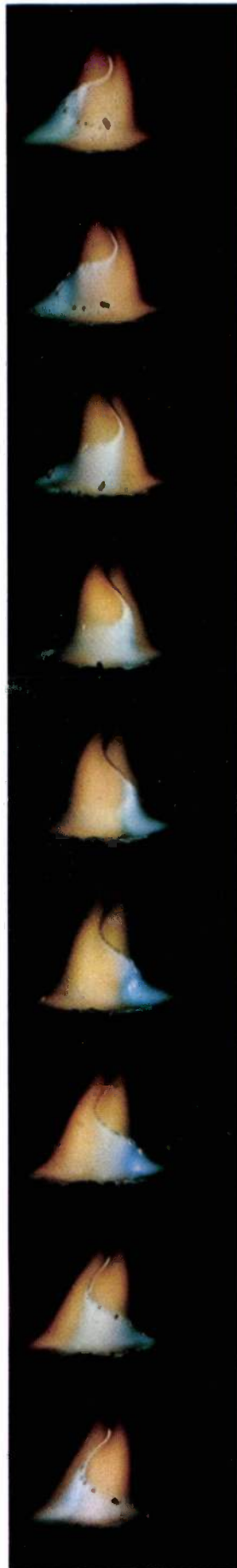
Fig. 3. Photograph of the plasma-MIG process with a thin stationary arc. The brighter arc between the tip of the filler wire and the workpiece can be seen at the centre of the plasma arc.

average diameter of 0.9 mm, travelling at an approximate velocity of 2.2 m/s. The high energy density in the arc causes the molten zone to penetrate deeply into the workpiece, producing welds like those in *fig. 4* (see page 24). The stationary type of arc can also be used for high-speed welding of sheet steel: 1 mm stainless steel sheet can be welded at a speed of 10 m/min.

Fig. 5a is a photograph of the rotating arc obtained when the critical current is exceeded. The fluid end of the wire assumes the form of a conical helix and rotates very rapidly within the almost unchanged plasma arc. The arc extends from the fluid wire tip to the workpiece and describes a vertical cylindrical surface as it rotates. Here this surface has a diameter of 8 mm. Fig. 5b gives a clear picture of the rotating fluid part of the wire and of the droplets detaching from it. High-speed films like this — in this case taken at a speed of 4000 frames per second — enable useful quantitative data to be derived. For example, the frequency of rotation at a given wire thickness is found to be related to the current in the filler wire and to lie between 130 and 270 revolutions per second. Another example relates to the transfer of the filler metal. If the conditions are the same as those in the stationary process illustrated in fig. 3, apart from the current in the wire and the wire feed, then at least 3000 droplets with a mean diameter of 0.6 mm

a   *b*

Fig. 5. *a*) Photograph of the plasma-MIG process with rotating arc. The contours of the surface of the molten wire tip can be seen inside the plasma. *b*) Series of frames from a high-speed film (4000 frames per second) of the rotating arc.
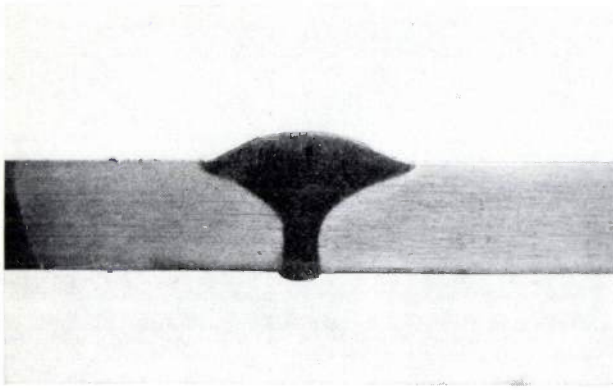
Fig. 4. Square-butt weld in plate of 8 mm stainless steel welded by the plasma-MIG method using a thin arc. The filler wire (PZ 6061) had a diameter of 1.2 mm and was also of stainless steel. The plasma gas was argon (6.5 l/min), the shielding gas a mixture of argon (10 l/min) and CO_2 (5 l/min). The diameter of the plasma exit was 6 mm, and its distance to the workpiece was 15 mm. Plasma-arc current 100 A at 50 V. Current in the filler wire 180 A at 40 V. Both electrodes were positive with respect to the workpiece. The rate of travel was 40 cm/min at a deposition rate of 95 grams of filler metal per minute. Under these conditions the critical current through the wire at which the arc starts to rotate is approximately 250 A.

Fig. 6. Weld bead of stainless steel (filler wire PZ 6061, diameter 1.2 mm) on a 10 mm steel plate. Plasma gas 6.5 l argon per minute, shielding gas 12 l argon + 3 l CO_2 per minute. Nozzle diameter 10 mm, distance from torch to workpiece 22 mm. Plasma-arc current 100 A at 45 V. Current in the filler wire 400 A at 40 V, both electrodes are positive with respect to the workpiece. Travel rate 23 cm/min, deposition rate 300 g/min, bead width 40 mm.

are transferred per second. The detached droplets move in an almost vertical path to the workpiece, with a velocity of about 2.6 m/s.

With the rotating arc there is no question of the deep penetration shown in fig. 4, since the electrical and thermal energy is spread over a larger surface of the workpiece. In this case a relatively wide weld bead is formed (in *fig. 6* nearly 40 mm) with shallow penetration in the metal sheet.

This welding method appears to be very suitable for cladding chemical storage tanks and nuclear reactor vessels with a corrosion-resisting layer of stainless steel. The high deposition rate is an economic advantage in such work. With a single welding torch a deposition rate of 30 kilograms per hour has already been obtained in our laboratory. This rate is high compared with that obtainable with other methods.

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

W. G. Essers, G. Jelmorini and Dr Ir G. W. Tichelaar are with Philips Research Laboratories, Eindhoven.

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, Netherlands   *E*

Mullard Research Laboratories, Redhill (Surrey), England   *M*

Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes (Val-de-Marne), France   *L*

Philips Forschungslaboratorium Aachen GmbH, Weißhausstraße, 51 Aachen, Germany   *A*

Philips Forschungslaboratorium Hamburg GmbH, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany   *H*

MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium.   *B*

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, Netherlands.

V. Belevitch: Dissipative filters.
Aspects of network and system theory, publ. Holt-Rinehart-Winston, New York 1971, pp. 267-285.   *B*

G. Bergmann: Polarisationseigenschaften von Glühlicht.
Math. naturwiss. Unterricht **25**, 133-139, 1972 (No. 3).   *A*

J. Bloem & A. H. Goemans (Philips Semiconductor Development Laboratory, Nijmegen): Slip in silicon epitaxy.
J. appl. Phys. **43**, 1281-1283, 1972 (No. 3).

J. van den Boomgaard: Stability of the high-temperature phase $FeSi_2$.
J. Iron and Steel Inst. **210**, 276-279, 1972 (No. 4).   *E*

L. Boonstra & F. L. J. Sangster: Analog functions fit neatly onto charge transport chips.
Electronics **45**, No. 5, 64-71, Feb. 28, 1972.   *E*

J. Borne: Analysis of the educational system.
Acta Electronica **14**, 289-293, 1971 (No. 3). (*Also in French, pp. 283-287.*)   *L*

C. J. Bouwkamp: Note on an asymptotic expansion.
Indiana Univ. Math. J. **21**, 547-549, 1971 (No. 6).   *E*

J. R. Brandsma & B. L. A. Waumans: A common bus switch.
Proc. Int. Computing Symp., Venice 1972, pp. 446-454.   *E*

P. Branquart, J. Lewi, M. Sintzoff & P. L. Wodon: The composition of semantics in ALGOL 68.
Comm. ACM **14**, 697-708, 1971 (No. 11).   *B*

C. H. J. van den Brekel & P. J. Severin: Control of the deposition of silicon nitride layers by 2537 Å radiation.
J. Electrochem. Soc. **119**, 372-376, 1972 (No. 3).   *E*

E. Bruninx: The evaluation of errors caused by flux inhomogeneities in thermal neutron activation analysis.
Anal. chim. Acta **60**, 207-217, 1972 (No. 1).   *E*

K. H. J. Buschow & H. J. van Daal: Comparison of anomalies observed in U- and Ce-intermetallics.
AIP Conf. Proc. **5**, 1464-1477, 1971 (No. 2).   *E*

M. Cazabat, C. Schiller, D. Diguet (R.T.C. La Radiotechnique Compelec, Caen) **& J. C. Lefèvre** (R.T.C. etc., Caen): Etude par microscopie à balayage des corrélations entre images de cathodoluminescence et de conduction des jonctions dans GaAs.
A.V.I.SEM 71, Coll. int. Versailles 1971 (Suppl. Le Vide No. 152), pp. 401-406.   *L*

T. Chisholm, J. Hubregtse (Groningen University), **G. F. Weston & E. E. Windsor**: The use of a channel electron multiplier with a residual gas analyser.
Residual gases in electron tubes, Proc. 4th int. Conf., Florence 1971, pp. 165-173; 1972.   *M*

J. B. Clegg: Reduction in secondary ion fogging on photographic plates used in mass spectrography.
Anal. Chem. **44**, 1100-1101, 1972 (No. 6).   *M*

J. Cornet & D. Rossier: Technique d'évaporation et caractérisation physique de couches minces amorphes de chalcogénures semi-conducteurs.
A.V.I.SEM 71, Coll. int. Versailles 1971 (Suppl. Le Vide No. 152), pp. 443-453.   *L*

C. Crevecoeur & H. J. de Wit: The preparation of α-As_2Se_3 crystals.
J. Crystal Growth **12**, 334-336, 1972 (No. 4).   *E*

P. Delsarte: Bounds for unrestricted codes, by linear programming.
Philips Res. Repts. **27**, 272-289, 1972 (No. 3).   *B*

P. Delsarte, J. M. Goethals & J. J. Seidel (Eindhoven University of Technology): Orthogonal matrices with zero diagonal, II.
Canad. J. Math. **23**, 816-832, 1971 (No. 5).   *B*

P. A. Devijver: A general order Minkowski metric pattern classifier.
AGARD Conf. Proc. No. 94, 18.1-18.11, 1971.   *B*

A. M. van Diepen, K. H. J. Buschow & J. S. van Wieringen: Study of the Mössbauer effect, magnetization, and crystal structures of the pseudobinary compounds $ThCo_{5-5x}Fe_{5x}$ and $ThNi_{5-5x}Fe_{5x}$.
J. appl. Phys. **43**, 645-650, 1972 (No. 2). *E*

A. M. van Diepen & R. P. van Stapele: Mössbauer effect of ^{57}Fe in cubic $Cd_{0.98}Fe_{0.02}Cr_2S_4$.
Phys. Rev. B **5**, 2462-2466, 1972 (No. 7). *E*

H. Dötsch: Magnetoelastic YIG delay lines with linear dispersion.
J. appl. Phys. **43**, 1923-1927, 1972 (No. 4). *H*

H. C. van den Elzen & P. van der Wurf: A simple method of calculating the characteristics of FSK signals with modulation index 0.5.
IEEE Trans. COM-20, 139-147, 1972 (No. 2). *E*

P. van Engelen, W. Boon & J. Dieleman: Electron nuclear double resonance of cobalt doped cadmium sulphide.
J. Phys. Chem. Solids **33**, 1041-1050, 1972 (No. 5). *E*

P. van Engelen & R. P. van Stapele: ENDOR of Co^{2+} in CdS.
J. magn. Resonance **7**, 122-123, 1972 (No. 2). *E*

M. Favreau (Thomson-CSF), **R. Genève, L. Goussot** (ORTF), **S. Lacharnay** (ORTF), **A. Laurens** (TRT) & **J. Polonsky** (Thomson-CSF): La radiodiffusion sonore et visuelle.
Onde électr. **51**, 594-604, 1971 (No. 7). *L*

N. E. Goddard: Instantaneous frequency-measuring receivers.
IEEE Trans. MTT-20, 292-293, 1972 (No. 4). *M*

J. J. Goedbloed: Determination of the intrinsic response time of semiconductor avalanches from microwave measurements.
Solid-State Electronics **15**, 635-647, 1972 (No. 6). *E*

G. Groh, G. S. Hayat (SUNY, Stony Brook, N.Y., U.S.A.) & **G. W. Stroke** (SUNY, Stony Brook): X-ray and γ-ray imaging with multiple-pinhole cameras using a posteriori image synthesis.
Appl. Optics **11**, 931-933, 1972 (No. 4). *H*

P. Guétin & G. Schréder: Band-structure effects in metal-GaSb tunnel contacts under pressure.
Phys. Rev. Letters **27**, 326-330, 1971 (No. 6). *L*

P. Guétin & G. Schréder: Influence d'un bombardement ionique de faible énergie sur les caractéristiques des barrières Schottky.
A.V.I.SEM 71, Coll. int. Versailles 1971 (Suppl. Le Vide No. 152), pp. 80-85. *L*

G. J. van Gurp: Precipitation of Si in evaporated Al/Si films.
Scripta metall. **6**, 361-366, 1972 (No. 5). *E*

P. Hansen: Ferromagnetic resonance in ruthenium-doped gadolinium iron garnet.
Phys. Rev. B **5**, 3737-3746, 1972 (No. 9). *H*

K. H. Härdtl & D. Hennings: Distribution of *A*-site and *B*-site vacancies in (Pb,La)(Ti,Zr)O_3 ceramics.
J. Amer. Ceramic Soc. **55**, 230-231, 1972 (No. 5). *A*

H. Haug & K. Weiss: Reduced density-matrix equations and superfluid hydrodynamics.
Physica **59**, 29-46, 1972 (No. 1). *E*

E. E. Havinga (III, IV), **H. Damsma** (III, IV) & **J. M. Kanis** (IV): Compounds and pseudo-binary alloys with the $CuAl_2$(*C*16)-type structure, III. Stability and competitive structures, IV. Superconductivity.
J. less-common Met. **27**, 269-280 & 281-291, 1972 (No. 3). *E*

J. H. C. van Heuven & T. E. Rozzi: The invariance properties of a multivalue *n*-port in a linear embedding.
IEEE Trans. CT-19, 176-183, 1972 (No. 2). *E*

K. R. Hofmann: Stability theory for thin Gunn diodes with dielectric surface loading.
Electronics Letters **8**, 124-125, 1972 (No. 5). *E*

K. R. Hofmann & H. 't Lam: Suppression of Gunn-domain oscillations in thin GaAs diodes with dielectric surface loading.
Electronics Letters **8**, 122-124, 1972 (No. 5). *E*

E. P. Honig: Theory and measurement of the suspension effect.
J. electroanal. Chem. interf. Electrochem. **37**, 249-266, 1972. *E*

F. A. de Jonge & W. F. Druyvesteyn: Calculations and experiments related to the magnetostatics of bubble domains.
Festkörperprobleme **12**, 531-597, 1972. *E*

B. A. Joyce & J. H. Neave: Reply to comments on the interaction of oxygen with Si(111) surfaces.
Surface Sci. **30**, 710-711, 1972 (No. 3). *M*

H.-G. Junginger & W. van Haeringen: Calculation of three-dimensional refractive-index field using phase integrals.
Optics Comm. **5**, 1-4, 1972 (No. 1). *A, E*

D. Kasperkovitz: An integrated 11 MHz *p-n-p-n* shift register.
Solid-State Electronics **15**, 501-504, 1972 (No. 5). *E*

A. Klopfer: Recent studies on electronic desorption from metals used in vacuum systems.
J. Vacuum Sci. Technol. **9**, 301, 1972 (No. 1). *A*

M. Kock & G. Rabe (Elektro Spezial, Bremen): Parallel and time sequential optical multiplex systems for pattern recognition.
Optics Comm. **5**, 73-77, 1972 (No. 2). *H*

G. Landvogt: A novel frequency analog force transducer with vibrating plate.
VDI-Berichte No. 176, 99-105, 1972. *H*

S. Leblanc: Enseignement assisté par ordinateur pour le recyclage des techniciens.
Acta Electronica **14**, 379-399, 1971 (No. 4). *L*

P. E. Leuthold (Federal Institute of Technology, Zürich): A new concept for the realization of data modems with integrated digital filters and modulators.
Philips Res. Repts. **27**, 223-243, 1972 (No. 3).

R. Loppes, R. Matagne (both with Université de Liège) & P. J. Strijkert: Complementation at the *arg-7* locus in Chlamydomonas reinhardi.
Heredity **28**, 239-251, 1972 (No. 2).   *E*

R. Loppes (Université de Liège) & P. J. Strijkert: Arginine metabolism in Chlamydomonas reinhardi.
Mol. gen. Genet. **116**, 248-257, 1972 (No. 3).   *E*

M. Martin: Analyseur de réponses pour expérimentations pédagogiques.
Acta Electronica **14**, 341-359, 1971 (No. 4).   *L*

D. Mateika: Growth of MnTe single crystals from nonstoichiometric melts by liquid encapsulation.
J. Crystal Growth **13/14**, 698-701, 1972.   *H*

L. J. Meuleman: Injection frequency locking of the avalanche transit-time oscillator.
Philips Res. Repts. **27**, 201-222, 1972 (No. 3).   *E*

R. F. Mitchell: Generation and detection of sound by distributed piezoelectric sources.
Thesis, London 1971.   *M*

F. L. van Nes (Institute for Perception Research (IPO), Eindhoven): Determining temporal differences with analogue and digital time displays.
Ergonomics **15**, 73-79, 1972 (No. 1).

H. W. Newkirk & A. J. R. de Kock: Crystal perfection of tetraphenyl tin and its isomorphs.
Phys. Stat. sol. (a) **11**, K 101-102, 1972 (No. 2).   *A, E*

K. H. Nicholas, B. J. Goldsmith, J. H. Freeman (AERE, Harwell), G. A. Gard (AERE), J. Stephen (AERE) & B. J. Smith (AERE): The evaluation of the Harwell-Lintott industrial ion implantation machine by making silicon planar resistors.
J. Physics E **5**, 309-310, 1972 (No. 4).   *M*

A. van Oostrom: Requirements for partial pressure analysers from the user's point of view.
Vacuum **22**, 15-17, 1972 (No. 1).   *E*

A. E. Pannenborg (Philips Board of Management, Eindhoven): Solid state physics.
Indian & Eastern Engr. Nov. 1971, 527-533.

D. H. Paxman, R. J. Tree & C. E. C. Wood: Unstable domains in solution-grown epitaxial InP.
Electronics Letters **8**, 241-243, 1972 (No. 9).   *M*

C. Piaget & Ph. Jarry: Applications des semiconducteurs à la photoémission et à l'émission secondaire.
A.V.I.SEM 71, Coll. int. Versailles 1971 (Suppl. Le Vide No. 152), pp. 427-442.   *L*

Ph. Piret: Structure and error probability of burst-correcting convolutional codes.
Philips Res. Repts. **27**, 244-256, 1972 (No. 3).   *B*

Ph. Piret: Convolutional codes and irreducible ideals.
Philips Res. Repts. **27**, 257-271, 1972 (No. 3).   *B*

J. E. Ralph & J. M. Woodcock: A new filamentary model for voltage formed amorphous oxide films.
J. non-cryst. Solids **7**, 236-250, 1972 (No. 3).   *M*

H. Rau: Apparatus for measurement of high vapor pressures and densities at high temperatures under isothermal conditions.
Rev. sci. Instr. **43**, 831-832, 1972 (No. 5).   *A*

K. Rennicke: Drehzahlregelung mit einphasig gestellten Induktionsmotoren.
Elektrotechn. Z. B **24**, 222-225, 1972 (No. 9).   *H*

F. Rondelez & H. Arnould: Déformations de la texture planaire d'un cholestérique à grand pas sous l'action d'un champ électrique.
C.R. Acad. Sci. Paris **273B**, 549-552, 1971 (No. 13).   *L*

D. J. Schipper & Th. W. Lathouwers: Bromine incorporation in sodalite.
Inorg. nucl. Chem. Letters **8**, 449-452, 1972 (No. 5).   *E*

P. J. Severin: Interpretation of the infrared thickness measurement of epitaxial layers.
Appl. Optics **11**, 691-692, 1972 (No. 3).   *E*

J. A. G. Slatter: Fundamental modelling of cylindrical-geometry bipolar transistors.
Electronics Letters **8**, 222-223, 1972 (No. 9).   *M*

M. J. Sparnaay: Ion-size corrections of the Poisson-Boltzmann equation.
J. electroanal. Chem. interf. Electrochem. **37**, 65-70, 1972.   *E*

W. T. Stacy & W. Tolksdorf: Growth-induced magnetic anisotropy in yttrium iron garnet.
AIP Conf. Proc. **5**, 185-189, 1971 (No. 1).   *E, H*

A. Stegherr: Korrosion von hochdispersen, aus Eisen-, Chrom- und Kobaltsalzlösungen reduzierten Metallpulvern.
Werkstoffe und Korrosion **23**, 180-186, 1972 (No. 3).   *A*

B. D. H. Tellegen: Stability of negative resistances.
Int. J. Electronics **32**, 681-686, 1972 (No. 6).   *E*

M. J. J. Theunissen: Etch channel formation during anodic dissolution of n-type silicon in aqueous hydrofluoric acid.
J. Electrochem. Soc. **119**, 351-360, 1972 (No. 3).   *E*

J.-P. Thiran: Recursive digital filters with maximally flat group delay.
IEEE Trans. **CT-18**, 659-664, 1971 (No. 6).   *B*

J.-P. Thiran: Equal-ripple delay recursive digital filters.
IEEE Trans. **CT-18**, 664-669, 1971 (No. 6).   *B*

W. Tolksdorf & F. Welz: The effect of local cooling and accelerated crucible rotation on the quality of garnet crystals.
J. Crystal Growth **13/14**, 566-570, 1972.   *H*

J. van der Veen, W. H. de Jeu, A. H. Grobben & J. Boven: Low melting liquid crystalline p,p'-di-n-alkyl-azoxy- and azobenzenes.
Mol. Cryst. liq. Cryst. **17**, 291-301, 1972 (No. 3/4).   *E*

M. L. Verheijke: A computer program in ALGOL-60 for the location and evaluation of peaks in Ge(Li) gamma ray spectra.
J. radioanal. Chem. **10**, 299-314, 1972 (No. 2).   *E*

Q. H. F. Vrehen & A. J. Breimer: Spectral properties of a pulsed dye laser with monochromatic injection. Optics Comm. 4, 416-420, 1972 (No. 6).   *E*

F. W. de Vrijer: Proposal for an integrated videophone signal.
Proc. Int. Zürich Seminar on Integrated Systems for Speech, Video and Data Communications, 1972, pp. F 7 (1)-(5).   *E*

K. Weiss: Doppelschicht und Gleichstrompolarisation von AgBr.
Ber. Bunsen-Ges. phys. Chemie 76, 379-384, 1972 (No. 5).   *E*

H. W. Werner, H. A. M. de Grefte & J. v.d. Berg: The measurement of small ion currents with the aid of photomultipliers.
Int. J. Mass Spectrom. Ion Phys. 8, 459-474, 1972 (No. 5).   *E*

H. J. de Wit: Hall effect of an inhomogeneous material. J. appl. Phys. 43, 908-913, 1972 (No. 3).   *E*

S. Wittekoek & D. E. Lacklison: Investigation of the origin of the anomalous Faraday rotation of $Bi_xCa_{3-x}Fe_{3.5+0.5x}V_{1.5-0.5x}O_{12}$ by means of the magneto-optical Kerr effect.
Phys. Rev. Letters 28, 740-743, 1972 (No. 12).   *E, M*

J. Wolter: Experimental study of nonlinear interaction between Bleustein-Gulyaev waves in a piezoelectric plate and an adjacent semiconductor.
Physics Letters 38A, 479-480, 1972 (No. 7).   *E*

C. E. C. Wood, R. J. Tree & D. H. Paxman: Solution-grown epitaxial InP for high-efficiency circuit-controlled microwave oscillators.
Electronics Letters 8, 171-172, 1972 (No. 7).   *M*

Contents of Philips Telecommunication Review 30, No. 4, 1972:

J. Mulder: 1 kW all-solid-state HF communications transmitter RZ 500 (pp. 137-146).
J. P. A. Aarts: The professional communications receiver RO 150 (pp. 147-155).
J. P. A. Aarts: PARCS, Automatic Remote Control System for HF radio stations (pp. 156-166).
M. M. Jung & J. de Boer: Comparison of the traffic handling capacity of subscriber transpositions (pp. 167-172).
R. Bodart: Telegraph distortion introduced by TDM systems (pp. 173-181).
W. Stimpel: TV transmission equipment at Munich Olympics (p. 182).

Contents of Electronic Applications Bulletin 31, No. 4, 1972:

E. Xanthoulis: Ten-bit digital-to-analog converter with thin-film ladder network (pp. 207-217).
J. Mulder: Large-signal behaviour of r.f. power transistors, Part 2. Computer program (pp. 218-233).
L. J. G. Berière: Measurement of a.m. suppression in an intercarrier sound channel (pp. 234-243).
R. W. Cooper: Infra-red intensity modulators using Faraday rotation in YIG (pp. 244-257).

Contents of Mullard Technical Communications 12, No. 116, 1972:

M. Michaelides: Microstrip transmission lines: impedance matching, coupling and filtering (pp. 170-196):
   Tapered transmission lines for impedance matching sections (pp. 170-176);
   Microstrip directional couplers (pp. 177-187);
   Interdigital filters and experimental microstrip filter design at 470 MHz (pp. 188-192);
   Microstrip bandpass filter for 0.7 to 1.4 GHz (pp. 193-196).
J. Gerritsen: A new method of colour purity adjustment (pp. 197-200).

Contents of Valvo Berichte 17, No. 1, 1972:

W. Kühl: Fernsehaufnahmeröhren für niedrige Strahlungsleistungen in LLL TV-Systemen (pp. 1-16).
D. Schwabe & J. Wüstehube: Digitale Teiler mit integrierten Schaltungen zur Untersetzung von Impulsfolgen (pp. 17-42).
W. Golombek: Verzerrungen elektronischer Verstärker (pp. 43-60).

A water-pressure operated control system for dishwashers

P. L. Holster, C. J. Th. Potters and H. F. G. Smulders

*In the automatic washing machines at present on the market the various washing pro-
grammes are controlled by an electromechanical system. However, by using 'fluid-logic'
elements it is now possible to make compact control systems that work entirely from the
water pressure. The experimental system for dishwasher control described in this article
is one of the results of investigations at Philips Research Laboratories.*

Introduction

In an automatic washing machine the various parts
of the preselected programme (prewash, main wash,
rinse, etc.) have to be worked through automatically
in the correct sequence. Each part of the programme
requires in turn a number of successive operations
from the machine; for the main wash, for example,
these could be: filling to the high level while adding a
detergent, heating to 60 °C, washing and finally drain-
ing off the water. To make this whole process take place
automatically a fairly complicated control system is
needed.

For many of the machine operations the only pro-
gramming required is to set the operating time. A
control system for washing machines therefore always
contains a timer, consisting of a miniature synchronous
motor, or stepping motor, which drives a camshaft on
which cams operate switches controlling the various
parts of the machine. The whole system is generally
referred to as a programme timer. There are some
machine operations, however, where the most important
factor is not the operating time but the attainment of
a certain final state: filling to a specified level, heating
to a specified temperature. During these phases of the
process the programme timer therefore has to be
stopped until the final state is reached. This makes
the control system more complicated, especially where
some conditions have to be satisfied simultaneously.
For example, the heating may only be switched on if
the programme timer gives the command, *and* if the
drum has been filled to a specified level, *and* if

the final temperature has not yet been reached; the
drum must be emptied when the programme requires,
or if a fault in the system makes the machine liable to
overflow. Such combinations of conditions can be
described with logic functions, in this case AND and
OR functions. To meet all these conditions, the control
system of a washing machine must therefore contain a
logic circuit as well as a programme timer. This logic
circuit controls the operation of the machine and is
controlled by signals from sensors that detect quanti-
ties such as level or temperature. In conventional
control systems all these components are electrical or
electromechanical.

A few years ago E. A. Muijderman of Philips Re-
search Laboratories proposed the idea of developing
a control system for washing machines that would
work entirely from the water pressure, making use of
the new technique of fluid logic or 'fluidics' [1]. Al-
though a water-operated logic circuit could also be
built with conventional hydraulic techniques, the result-
ing systems would be too complicated and too bulky
for this application. The camshaft drive is no problem;
there is sufficient energy available in the water mains
to allow a small hydraulic motor to be used for this
purpose. (The Dutch water-supply authority, for
example, officially guarantees a water pressure of
1 atmosphere at the tap, and in practice it is usually
much higher.)

*P. L. Holster, C. J. Th. Potters and H. F. G. Smulders are with
Philips Research Laboratories, Eindhoven.*

[1] Fluidics, or fluid logic, is the term used to describe the use of
fluids — gas or liquid — for carrying out logic or amplifica-
tion functions. Where a gas is used in the operation of fluid-
logic elements, the term gas logic or air logic is sometimes
used (in practice the gas is usually compressed air).

To investigate the possibilities of systems operating from the water pressure we have made an experimental control system for a dishwasher. We chose a dishwasher because it works with a much simpler programme than an ordinary washing machine. Before dealing with this system we shall first discuss the logic elements used.

The logic elements

In fluidics a distinction can be made between two groups of elements: those *with* and those *without* moving parts. The elements without moving parts work through the interaction of flowing fluids. We therefore call them *dynamic* elements. The other group includes a large number of devices in which components such as diaphragms, valves, foils or balls are displaced by the pressure of the fluid. This second group of elements can be further subdivided into *static* and *quasi-static* elements. In static elements there is no fluid flow except during switching from one state to another, while quasi-static elements allow fluid to flow in one of the states — though the quantity that flows is much smaller than with the dynamic elements.

The logic elements used in our control system are mainly quasi-static *diaphragm elements*, chosen because they are very simple and also reliable. While it might seem that the restrictors in these elements could easily become blocked, this was not found to happen in practice. Other static elements that we could have used instead would have required diaphragms of a type too complicated for the manufacture of a single prototype. The water consumption of the quasi-static elements is not a disadvantage in the control of a washing machine, since the water can be drained into the wash unit. All the elements employed are binary, and their input and output signals are either hydraulic or mechanical.

The input elements, which are required to convert mechanical signals (in this case the positions of cams) into water-pressure signals, are quasi-static elements with a valve, called *baffle nozzles*. The principle of this element is illustrated in *fig. 1*. The water supply is connected to the input A. When the valve (the baffle) B is closed, the pressure in the element, and hence the pressure at the output F, is equal to the input pressure. (We assume that no water is drawn off at the output, which implies that only the *pressure* is used as the output signal.) When the valve is opened, water flows through the element from the input A, through the restrictor R along the valve and to the outside. The restrictor is a constriction that acts as a flow resistance, and a pressure difference appears across it during flow.

As the valve opens and the flow through the element increases, the pressure difference becomes greater and the pressure in the element therefore falls. When the valve is fully open, the pressure in the element, and hence the output pressure, will be practically equal to the ambient pressure. The baffle nozzle is thus an analog device, the output pressure varying propor-
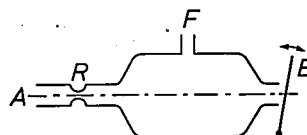
Fig. 1. Diagram illustrating the principle of the baffle nozzle. A input to which the water supply is connected. B valve (baffle) which opens and closes the outlet (the nozzle). When the valve is closed, the pressure at the output F is the same as at the input. When the valve is open, water flows through the element to the outside. Owing to the pressure difference across the restrictor R the output pressure falls nearly to the ambient pressure, while the input pressure remains high.

tionally with the position of the valve. In our system, however, only two positions of the valve are used: 'open' and 'closed', and the baffle nozzle therefore acts as a binary element.

Fig. 2a shows the baffle nozzle as designed for our control system. The valve B is opened against the force of the spring S when the valve rod is raised (by a cam on the camshaft of the programme timer). Briefly the device operates as follows: when the pressure at the input A is p_A, the output pressure p_F is equal to p_A when the valve is closed and is zero when the valve is open. (In this article we take p to be the excess pressure over the ambient pressure.) When the input pressure of the element is zero, the output pressure is always zero.

The baffle nozzle is usually used in such a way that the valve is opened when the machine has to perform a particular operation; the logic convention we have adopted is therefore: $B = 0$ means 'baffle closed', $B = 1$ means 'baffle open'. For the water pressure we have adopted a convention in which $A = 0$ means 'water pressure at A low', $A = 1$ means 'water pressure at A high'. With the aid of this description we can easily verify that the baffle nozzle operates as an AND gate with negation; F is only 1 if $A = 1$ *and* $B = 0$ (thus $B = not$ 1): $F = A.\bar{B}$. Fig. 2b shows the logic symbol for the element.

Fig. 3a gives a cross-section of the diaphragm element mostly used in our system. When the pressure at the input A of this element is high and the pressure at the input C is low, the pressure in the annular space that communicates with A will make the diaphragm flex downwards against the force of the spring S. A connection is then established between input A and

output F, and the output pressure will therefore be high. If high pressure is applied at both C and A, the pressure will then be the same on both sides of the diaphragm. The forces acting on the diaphragm are now approximately equal and the spring S is able to press the diaphragm against the seat, so that the connection between A and F is broken. Since water can flow away to the outside through the restrictor R, the output pressure is low. The restrictor can be omitted if the pressure drop is allowed to appear across a water-transmitting element connected to the output. If we regard the diaphragm device as a logic element, we see that it works as an AND gate with negation, the output pressure only being high if the pressure is high at A *and* low at C: $F = A.\bar{C}$ (fig. 3b).

We can also regard the diaphragm device in fig. 3 as a flow amplifier (again of course with negation), for if $C = 0$ a connection is established between A and F and the flow of water can be large, while the input C takes no water in either position. We make considerable use of the diaphragm element in our control system as a flow amplifier in series with a baffle nozzle. We do this because the baffle nozzle can only handle very small flows: when water is drawn off at the output, the output pressure drops even when the baffle is closed, owing to the pressure difference across the input restrictor.

Since quasi-static elements take water in the 'open' position, the 'high' pressure will gradually fall across a large number of these elements connected in series. In such an arrangement the 'low' pressure will not usually be constant either; even fully open valves will form small flow resistances that will have pressure differences across them, so that the low pressure will rise above the ambient pressure. The two pressure levels are therefore not defined as pressure *values* but as pressure *regions*. In the system that will be described here, the high level at a supply pressure p_s is defined by $p_s > p > 0.8\,p_s$ and the low level by $0.2\,p_s > p > 0$.

The control system has to operate a number of electrical switches (for the heating, the pump, etc.), and output elements are therefore required that can convert a water-pressure signal into a movement signal. This is done by a pair of static diaphragm elements; the simpler of the two is shown in *fig. 4a*. When the pressure at A is high, the diaphragm M pushes the rod F upwards against the force of the spring S (the space above the diaphragm is freely connected with the outside). The rod can operate a switch. If $F = 0$ denotes the rest position and $F = 1$ the upper position, then $A = 1$ corresponds to $F = 1$. We can describe this element by the simple relation $F = A$; it operates only as a signal transducer and does not perform a logic operation. Fig. 4b shows the symbol used for it.

The second output element used in our system, shown in *fig. 5a*, is somewhat more complicated. It has two inputs and two diaphragms; fixed between the diaphragms there is a disc D, of hard rubber. When pressure is applied to the space between the diaphragms
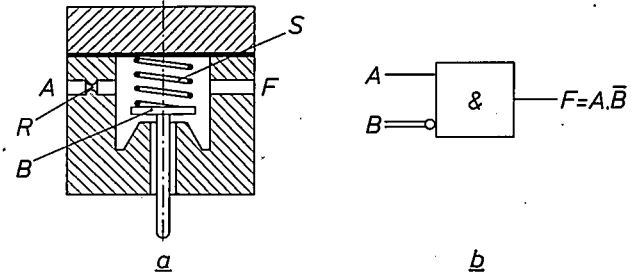

Fig. 2. *a*) The baffle nozzle used in the control system. B valve. S spring. The valve is opened by raising the valve rod; this is done by a cam on the camshaft of the programme timer. When the valve is closed ($B = 0$) a high input pressure ($A = 1$) corresponds to a high output pressure ($F = 1$). When the valve is open ($B = 1$) a high input pressure ($A = 1$) corresponds to a low output pressure ($F = 0$). If there is no input pressure ($A = 0$) the output pressure is always low ($F = 0$). *b*) Logic symbol for the baffle nozzle; a single line denotes a water-pressure signal, and a double line a mechanical signal. The element acts as an AND gate with negation, since $F = 1$ only if $A = 1$ *and* $B = 0$ (*not* 1).
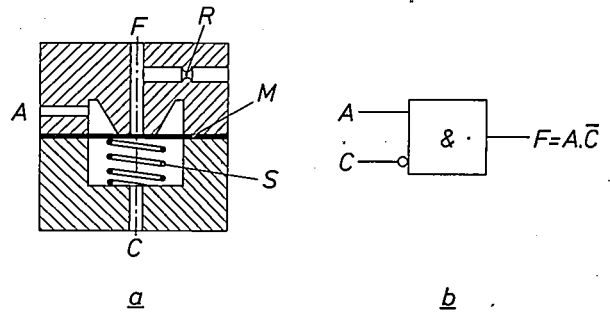

Fig. 3. *a*) The diaphragm element. A and C inputs, F output. When the pressure is high at A and low at C, the diaphragm M is pushed downwards so that a connection is established between A and F. The output pressure is then high. If the pressure is high at both A and C, the spring S can push the diaphragm upwards, breaking the connection. The pressure at F can now leak away through the restrictor R, which connects with the outside, so that in this situation the output pressure is low. *b*) Logic symbol for the diaphragm element. This element also acts as an AND gate with negation: $F = 1$ only if $A = 1$ *and* $C = 0$.
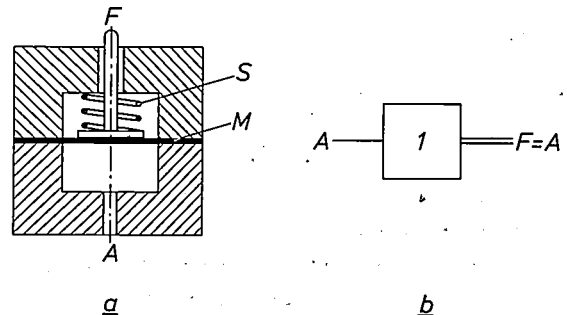

Fig. 4. *a*) Output element that converts a water-pressure signal into a movement (e.g. for operating an electrical switch). When the pressure at input A is high ($A = 1$), the rod F is at its highest position ($F = 1$). The operation of the element is thus represented by the relation $F = A$; it functions only as a signal converter. *b*) Logic symbol for this element.

via the input A, the upper diaphragm flexes upwards and pushes up the rod F. (The lower diaphragm flexes downwards at the same time, but this does not matter.) If pressure is applied to the input B and not A, then the lower diaphragm flexes upwards; the disc D transmits this movement to the upper diaphragm, which again pushes up the rod. The rod is of course also displaced when both A and B are under pressure. The element therefore operates as an OR gate: $F = 1$ when $A = 1$ *or* $B = 1$ (or when both are 1). The logic symbol for this element is shown in fig. 5*b*.

We have seen that both the element of fig. 3 and the output elements belong to the group of diaphragm elements. However, since the element of fig. 3 is by far the most widely used in our system, we shall reserve the term diaphragm element exclusively for this particular type, and where necessary indicate the output elements with their figure number.
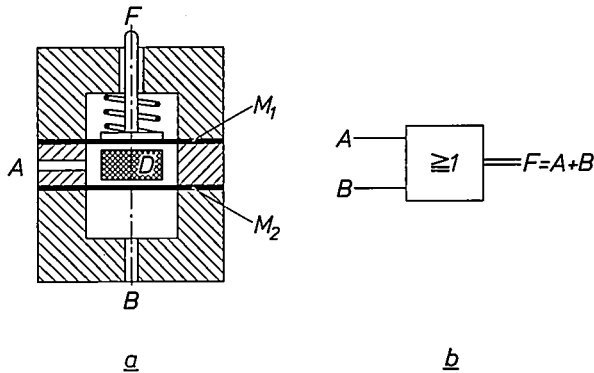
Fig. 5. *a*) Output element with two inputs A and B. Fixed between the two diaphragms M_1 and M_2 is a solid disc D (of rubber). The rod F is in its highest position when the pressure at A or at B is high, or when both pressures are high. The element thus acts as an OR gate: $F = 1$ if $A = 1$ *or* $B = 1$ (or if $A = 1$ and $B = 1$). *b*) Logic symbol for this element.

Operation of a dishwasher

In most dishwashers the dishes and other articles (the load) are arranged in racks and sprayed with jets of water from one or two spray arms which rotate because of the reaction from the jets of water. A predetermined quantity of water is taken for each washing or rinsing cycle. It is pumped to the spray arms from a sump at the bottom of the washing unit, drains back after spraying, is pumped up again, and continues to circulate in this way through the machine until the washing or rinsing cycle is completed. The detergent and the rinsing agent to prevent streaky drying can therefore be introduced into the machine at any appropriate point, and they automatically mix with the circulating water. The sump contains the heating element, and the used water is pumped away from the sump on the completion of a washing or rinsing cycle.

The water temperature is controlled by a thermostat, the water level by a 'pressostat' connected to the sump near the bottom.

In some machines the load is dried by blowing hot air through it, in others it is heated by the final rinse to a temperature at which the water still clinging to the load evaporates. This method is of course cheaper but takes longer.

In districts where the water is very hard, the machine has to be provided with a water softener. This works on the ion-exchange principle: the water is passed through a solid which replaces the Ca and Mg ions by Na ions. The Ca and Mg ions are retained in the softener, which therefore has to be regenerated occasionally with Na ions. This is done at the end of the programme by flushing the softener through with a solution of NaCl.
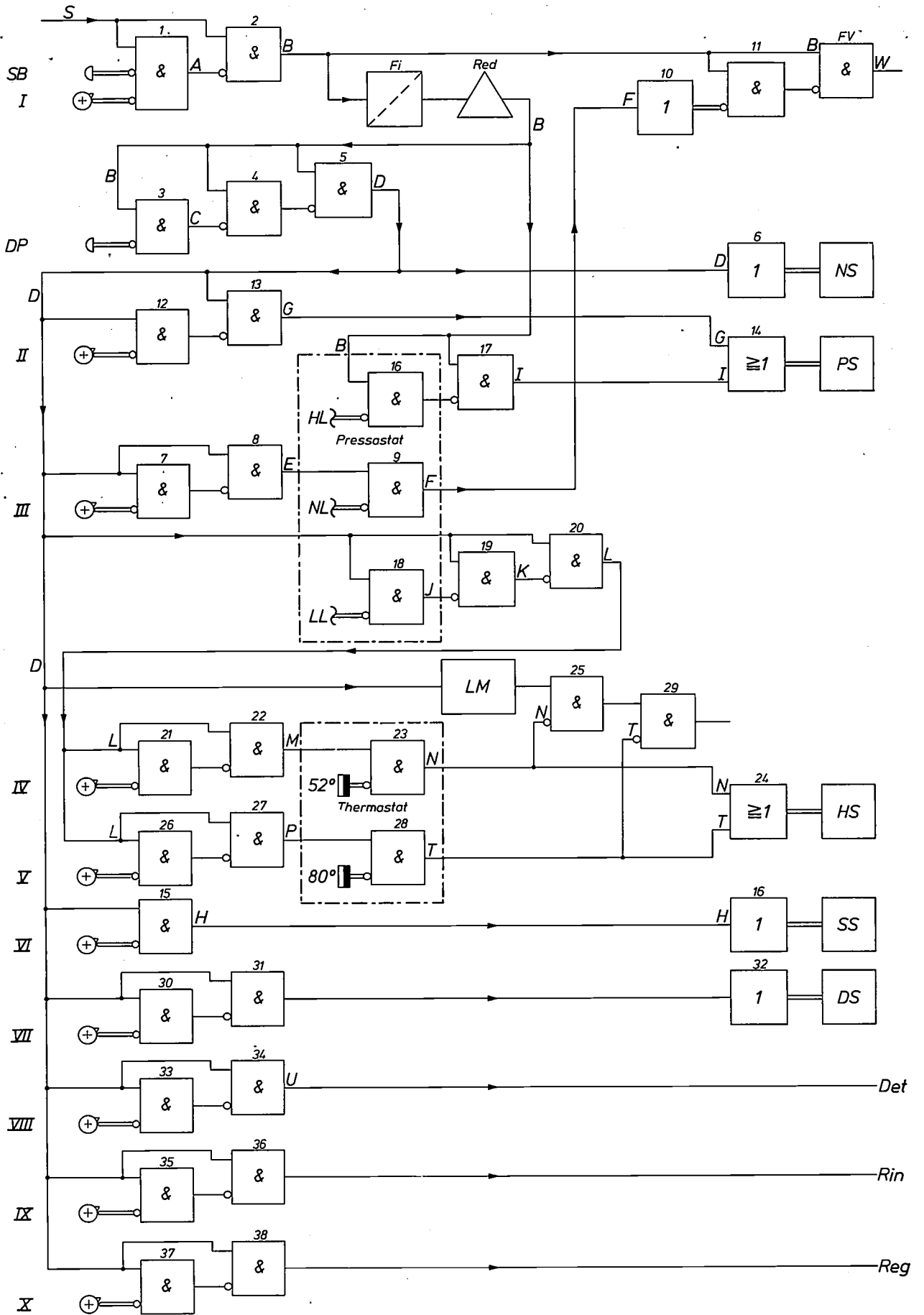
The control system

Our control system was designed for a dishwasher with a hot-air drier and water softener; the programme is shown in *fig. 6*.

The load is first rinsed with cold water, then washed, rinsed twice with cold water, then rinsed with hot water containing rinsing agent, and finally dried. During the drying the water softener is regenerated; since this inevitably leaves salt behind, the machine must first be rinsed clean before starting the programme. The figure shows the succession of machine operations during the various cycles of the programme. The valves of the input elements of the control system have to be opened and closed to this pattern. The operation of valve *VI* departs from the rule that an operation is always carried out when the appropriate valve is open;

Fig. 6. Schematic diagram of a programme for a dishwasher with hot-air drier and water softener. On the left is a list of the various operations of the machine with the numbers of the input valves *I* to *X* of the control system. An indication is given on the right of the sequence in which these valves must be opened as the machine works through the various parts of the programme (a dash means 'valve open'). The whole programme is completed in 40 steps of 1 minute each, plus the time needed for heating, during which the programme is interrupted.

Valve no.	Machine operations
I	Machine operating
II	Pump out
III	Fill
IV	Heat to 52 °C
V	Heat to 80 °C
VI	Spray (if valve shut)
VII	Dry
VIII	Add detergent
IX	Add rinsing agent
X	Regenerate

the spraying action is performed when the valve is closed. This is done to limit the water consumption of the control system, since the water is heated during the spraying operation, and the programme is not continued until the required temperature is reached, so that the spraying takes up by far the greatest proportion of the total time.

The valves are operated by cams on a shaft driven by a hydraulic motor. This is a linear motor in which a piston makes one stroke per minute; this movement is converted by a ratchet-wheel and pawl into a rotation of the camshaft. Since there are 40 positions in the programme (0 to 39 in fig. 6) the transmission is arranged in such a way that the camshaft is rotated 9° for every stroke of the piston, so that the whole programme is run through in one revolution of the camshaft. It takes 40 minutes to complete the programme, plus the total time needed to heat the water.

Fig. 7 shows the logic diagram of the control system. The input elements are shown on the left in the figure and the output elements, such as the electrical switches for operating the pump motors, are shown on the right. The logic elements are numbered 1 to 38. The connections between the elements are indicated with a single line if they relate to a fluid flow, and with a double line if the connection is a mechanical one. The letters denote input and output signals.

In setting 0 of the programme (see fig. 6) all the input valves are closed (except valve VI, of course). When the water supply is connected (S = 1) the output pressure of element 1 (a baffle nozzle with two valves) is high (A = 1), so that the diaphragm element 2 blocks the supply of water to the rest of the system (B = 0). The machine is switched on by pressing the pushbutton SB, causing the valve SB of element 1 to open. The

output pressure of element 1 is now low, and the diaphragm of element 2 opens the inlet (B = 1). The water can now flow in two directions, first to the filler valve FV (for the moment closed), and secondly via a filter Fi and a reducing valve Red to the door safety unit DP and the high-level part of the pressostat.

We shall first follow the path of the water through the door safety unit. This consists of a restrictor with a hose that allows water to flow when the door is open and is squeezed shut when the door is closed. This can be considered as a baffle-nozzle element (3); closing the door then corresponds to shutting the valve DP. When the water supply is opened (B = 1) and the door closed (DP = 0), the output pressure of element 3 is high (C = 1). In this situation the actual control part of the system must enter into operation. Since this requires a fairly large water flow, and element 3 cannot supply this, two diaphragm elements 4 and 5 are connected in series with this element as flow amplifiers. Element 5 passes the flow of water capable of supplying the control section (D = 1). (We have used two elements in this case since one element gives a negation, whereas here an input signal 1 must also result in an output signal 1.) A pressure now acts on the line that runs from top to bottom in fig. 7; the linear motor LM and a large number of input elements are connected to this line. The motor consequently starts up, causing valve I to open as soon as SB is operated and the door closed. The pushbutton can now be released and the machine left to complete the programme automatically. At the end of the programme the machine is stopped by valve I, which closes when the machine switches from setting 39 to setting 0, causing element 2 to interrupt the water supply. As long as the machine is in operation a neon lamp remains

0	Machine rinse			Prerinse				Wash						Rinse				Rinse				Hot rinse with rinsing agent									Dry and regenerate										0
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	0

alight; this is switched on by element 6 (an output element of the type in fig. 4) and the switch NS.

Let us now follow the programme step by step. In setting 1 valve III is opened to admit the water into the machine. The output pressure of baffle nozzle 7 is now low and the output pressure E of diaphragm element 8 is high. This pressure is applied to baffle nozzle 9, which forms part of the pressostat. In the rest position the valve of this element is closed; when subjected to the water pressure in the machine it opens as soon as the normal water level (NL) required for washing and rinsing is reached. As long as valve NL is closed (NL = 0) and valve III is open (E = 1) the output pressure of element 9 is high (F = 1). Elements 10 and 11 now cause the filler valve FV to open (W = 1), allowing water to flow into the machine. The filler valve does not close again until the pressure F is low, which is the case either when the normal water level is reached, so that valve NL opens, or when valve III closes. (The programme must of course be arranged so that for a given cycle time of the motor there is sufficient time for the required level to be reached before valve III closes again.)

In setting 2 of the programme, valve II is opened and the pump is started. Meanwhile the water continues to flow in, since the machine has first to be rinsed clean. When valve II is opened, element 13 gives the signal G = 1. This signal is applied to element 14 (an output element of the type in fig. 5) which operates the pump switch PS. In setting 3 of the programme valve III is again closed, stopping the inflow of water; pumping continues, and the water in the machine is completely drained away.

For the second part of the programme, the cold rinse, the filler valve is first opened to let the machine fill up to the normal level. Next, in setting 5 of the programme, the spray pump that sends the water to the spray arms is started. This is done by closing valve VI, which is open in the rest state. The output pressure of baffle nozzle 15 is then high (H = 1) and this signal causes element 16 to operate switch SS, which actuates the spray pump.

We shall not follow the programme further here, but will take a look at the remaining parts of fig. 7. In addition to the valve NL the pressostat contains two other valves: HL and LL. Valve HL (high level) is an overflow safety valve. In normal operation the valve stays closed, but it opens if a fault occurs, e.g. in the filler valve, that might allow the water level to rise too high. The element used is again a baffle nozzle. It is directly connected to the reducing valve so that, whatever the position of the door, if the water level is too high (HL = 1) the output pressure of element 16 is low and that of element 17 is high (I = 1). The pump switch is then operated via element 14.

The heating element must not of course be switched on before the machine has filled up far enough for the element to be completely submerged. This protection of the heating element is provided by the low-level valve LL of the pressostat. This is open when the machine is empty and closes as soon as the machine has filled to the lowest level at which the heating can safely be switched on. Baffle nozzle 18 then gives the signal J = 1, which causes diaphragm element 20 to open the inlet to baffle nozzles 21 and 26, allowing the heating to be switched on (L = 1). As in the door safety unit diaphragm elements 19 and 20 are connected as a flow amplifier without negation.

To heat the water to 52 °C valve IV is opened so that element 22 gives the signal M = 1. Element 23 is a baffle nozzle that forms part of the thermostat. When the water temperature is lower than 52 °C the valve of this element is closed, so that a high input pressure gives a high output pressure (N = 1) and the heating is switched on by means of element 24 and switch HS. The signal N is also applied to element 25. This shuts off the flow of water through the motor, which therefore stops and the programme is discontinued. (The reason why this element, like element 29, which performs the same function for heating to 80 °C, is located after the motor and not before it, will be explained when we deal with the motor.) When the water temperature has reached 52 °C the output pressure of element 23 goes to the low level, so that the heating is switched off, the motor starts up again, and the programme is resumed. The same functions are performed by valve V with elements 26-29 and the signals L, P and T for heating to 80 °C.

We are left with elements 30-38 to consider. When valve VII is opened (in setting 32 of the programme), elements 30, 31 and 32 actuate the switch DS of the hot-air drier. When valve VIII is opened, the output pressure of element 34 rises (U = 1). This water pressure is used to feed a quantity of detergent into the machine, e.g. by tilting a filler tray. In the same way, a signal from valve IX causes element 36 to add

Fig. 7. Logic scheme of the control system. The input elements are on the left: SB starting button; DP door safety unit; I to X valves of baffle nozzle elements, which are raised by cams on the camshaft. The output elements are on the right: FV filler valve; switches NS for a neon indicator lamp, PS for the pump, HS for the heating, SS for spraying and DS for drying; Det, Rin and Reg are outputs whose pressure is used for introducing the detergent, the rinsing agent and the regenerating salt. Also shown in the scheme are a filter Fi, a reducing valve Red, and the linear motor LM that drives the camshaft. The figures 1 to 38 denote the various logic elements, and the letters A to U denote pressure signals. The pressostat and the thermostat are surrounded by a dot-dash line.

rinsing agent, and finally, a signal from valve X causes element *38* to supply the regenerating salt to the water softener.

Design and construction

To make the construction of this type of control system as simple as possible, a design was used in which the whole system is assembled from a number of plastic plates separated by sheets of rubber — a kind of sandwich construction. The elements are introduced by milling holes in the plates; for a diaphragm element, for example, this can be a recess in one plate and a diaphragm seat in the next plate, so that the rubber sheet can serve as a diaphragm between these plates. At locations where there are no elements, the rubber sheets act as a seal. The connections be-

under the diaphragm. In practice the two elements therefore form a single space. We have located the restrictor R_1 of the baffle nozzle on the left of the diaphragm element, below the input A of the supply pressure, thus dispensing with the need for a separate supply line to the baffle nozzle, and making the design much simpler.

Before describing the construction of the complete system, we shall first discuss the design and operation of some components that have so far only been mentioned briefly.

The linear motor

The motor used for the intermittent drive of the camshaft in the programme timer is a linear hydraulic motor (see *fig. 9*). The motor piston MP gives one stroke per minute; this movement is controlled by the
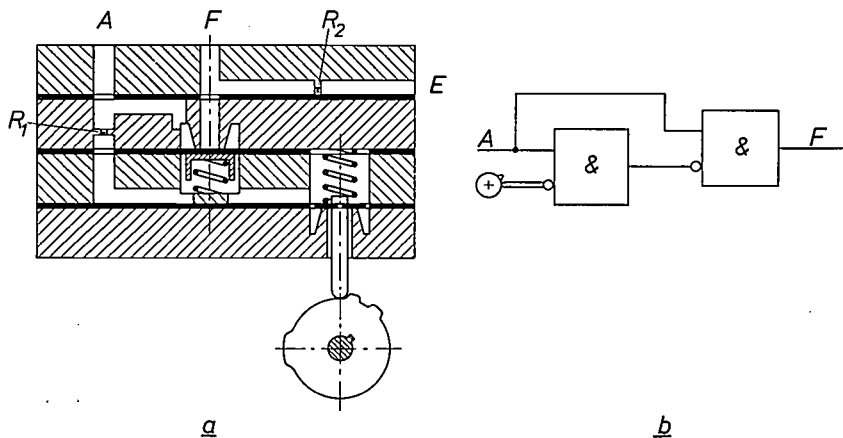
Fig. 8. *a*) Cross-section of the elements that convert the mechanical signal from a cam into a water-pressure signal for introducing the programme information. The elements are contained in a sandwich construction of plastic plates, separated by rubber sheets which act as a seal and also serve as diaphragms for the diaphragm elements. On the right is a baffle nozzle, whose valve is raised by a cam, and on the left can be seen a diaphragm element. A input for the supply pressure. F output. R_1 restrictor for the baffle nozzle. R_2 restrictor for the diaphragm element. E outlet. *b*) Logic diagram for this arrangement.

tween the elements are also milled into the plates, and where necessary holes are punched into the rubber sheets for this purpose. The other parts of the elements (springs, valves, etc.) are mounted in position when the plates and rubber sheets are assembled.

As an example of this sandwich construction *fig. 8a* shows a cross-section of part of the programme input, consisting of a baffle nozzle combined with a diaphragm element, which converts the position of a cam into a water-pressure signal. Fig. 8*b* gives the logic diagram for this arrangement. The elements in this design have the same form as in figures 2 and 3. Since the output signal of the baffle nozzle has to be applied to the control input of the diaphragm element, the inside of the baffle nozzle is connected with the space

valve V, and the cycle time is determined by means of the delay element D.

The control valve is located between two diaphragms in a space that freely communicates with the outside atmosphere. The position of the valve (upwards closed, downwards open) therefore depends only on the forces acting on these diaphragms. The space above the valve is divided into three by two circular seats. The water supply (supply pressure p_s) is connected to the annular space between the two seats; the outer space, where the pressure is shown as p_1, is connected via a restrictor with the space under the valve and above the rolling diaphragm MR of the piston (pressure p_2). The central space (pressure p_3) is connected with the delay element, consisting of the piston DP with rolling diaphragm

DR, and also with the 'blocking element' B, a kind of diaphragm element.

We shall now follow a single cycle of the motor, starting from the initial position in which the valve rests on the seats, all pressures are equal to zero and the motor piston is at its highest position. When the

rolling diaphragm, the other due to the spring. We assume that the spring force is constant over the whole piston stroke (it can be approximated fairly well by a highly prestressed weak spring), and we define a factor α such that the two forces acting on the piston balance one another when $p_2 = \alpha p_s$. As soon as p_2 rises to this

Fig. 9. Cross-section of the linear motor used for driving the camshaft. MP motor piston, MR rolling diaphragm providing the piston seal. V control valve. p_s supply pressure, p_e ambient pressure. p_1 pressure in the outer annular chamber above the control valve. p_2 pressure in the space under the valve and above the piston. D delay element with piston DP, sealed by rolling diaphragm DR. p_3 pressure above the piston DP. B blocking element. R_1, R_2 and R_3 restrictors.
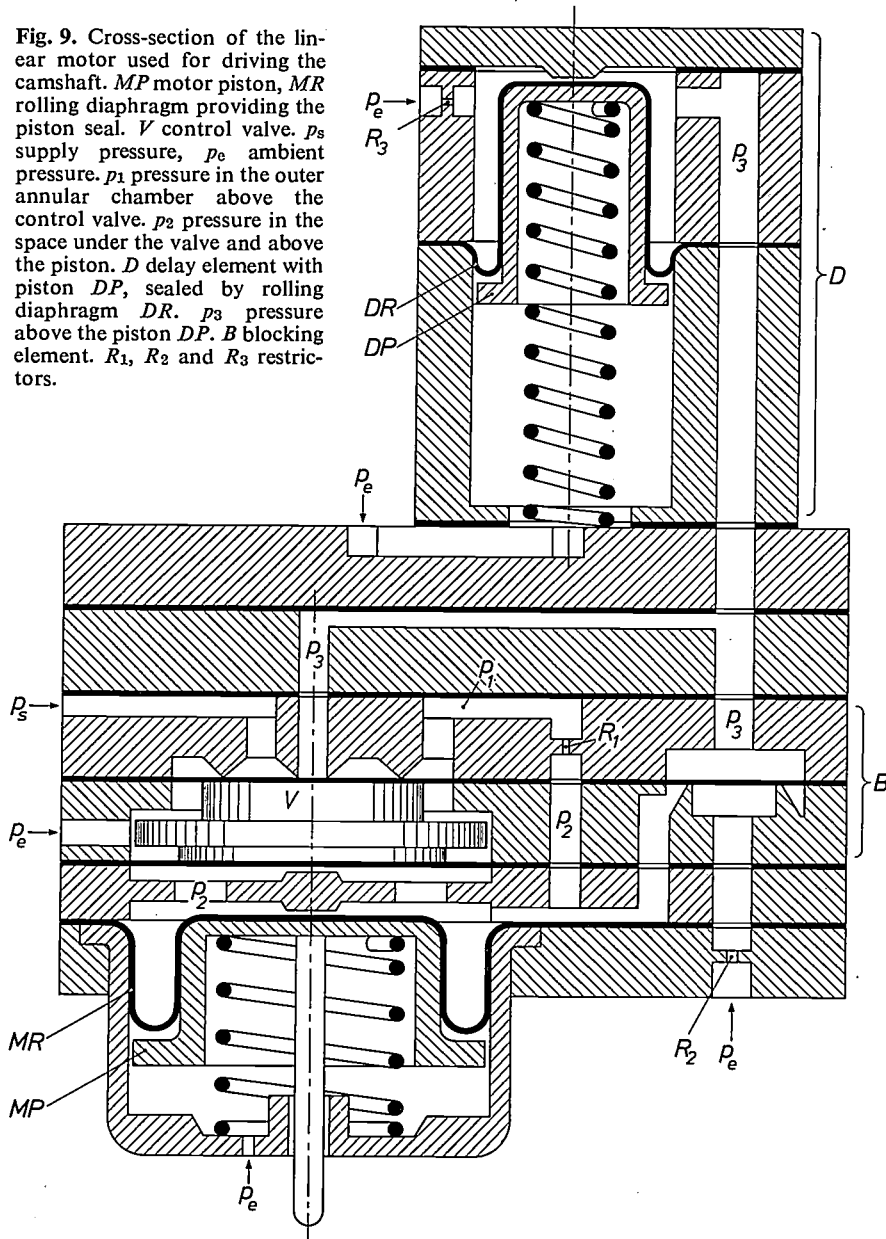
water supply is connected, the valve is pushed down, so that water starts to flow along the seats and the pressures p_1 and p_3 rise (see *fig. 10a*). The pressure p_1 rises almost to the value p_s. The pressure p_2 also rises, but owing to the pressure drop across the restrictor R_1 the value of p_2 remains lower than that of p_1. Moreover p_2 stops rising as soon as the piston begins to move, since neglecting inertial forces there are two forces acting on the piston, one due to the water pressure above the

value αp_s, the piston starts to move (at $t = t_2$, see fig. 10a and b). If the motor is not then under load, the magnitude of p_2 until the piston reaches its lowest position is determined solely by the opposing force exerted by the spring. This means that during the downward stroke p_2 cannot become greater than αp_s. Not until the piston reaches its lowest position (at $t = t_4$) can the pressure start to rise again.

The diaphragms above and below the control valve

are designed in such a way that a pressure $p_2 = \alpha p_s$ under the valve is not great enough to close the valve against the action of the pressures exerted from above (p_1; p_s and p_3). This does not happen until the pressure becomes higher when the piston reaches its lowest position. The water supply is then interrupted. The motor spring now tries to push the piston up again, but since both the valve and the blocking element are closed (the latter will be dealt with presently), the

cycle. The piston DP of the delay element is at its highest position. As soon as the control valve is opened and p_1 starts to rise, p_3 — the pressure above the rolling diaphragm DR of the delay element — also starts to rise. This increase in pressure also stops when the piston begins to move (at $t = t_1$, see fig. 10a and c). During this movement the equilibrium of forces is the same as was described for the motor piston; the pressure p_3 does not rise above a value βp_s, the factor β at
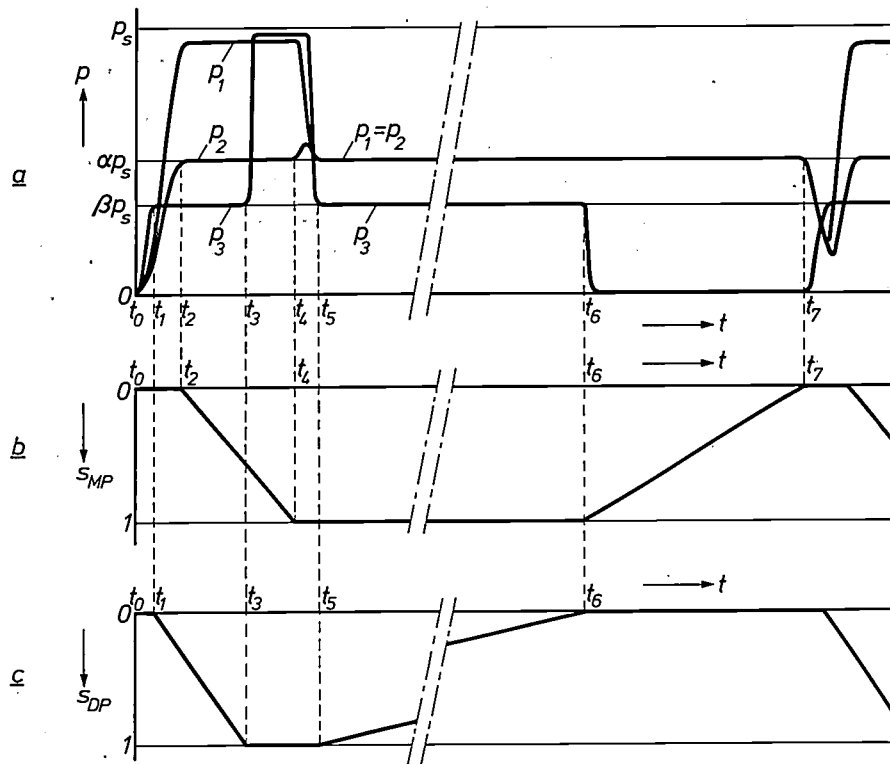
Fig. 10. The pressures p_1, p_2 and p_3 in the motor (a) and the positions s_{MP} of the motor piston and s_{DP} of the delay piston (b and c), plotted as functions of time. The positions of the pistons are given as fractions of their maximum stroke; the initial position (at $t = 0$) corresponds to the situation in fig. 9. The whole cycle of the motor is completed between t_0 and t_7. The downward stroke of the delay piston starts when p_3 is equal to βp_s, the downward stroke of the motor piston starts when p_2 is equal to αp_s. At the end of the downward stroke of the motor piston the control valve closes. The delay piston then starts on its upward stroke, keeping the blocking element closed. This blocks the upward stroke of the motor piston, which cannot start to move until the delay piston has reached its highest position, and opens the blocking element. The duration of the cycle completed by the motor is determined mainly by the duration of the upward stroke of the delay piston.

water above the rolling diaphragm cannot be displaced, and the piston therefore remains in its lowest position (see fig. 10b). In this situation the spring acts as a pressure source, so that we again have $p_2 = \alpha p_s$. Thus the pressure p_2 only rises above αp_s for a moment to close the valve, and then immediately returns to the value αp_s. Since no more water flows through R_1 after the valve has been closed, p_1 and p_2 are now equal, which means that on closing the valve the pressure p_1 also falls to the value αp_s ($t = t_5$ in fig. 10a).

To explain the operation of the delay element and the blocking element we return to the beginning of the

a given supply pressure — like α for the motor piston — being determined by the spring force and by the surface area of the rolling diaphragm. When the piston has reached its lowest position ($t = t_3$), p_3 rises to the supply pressure while the piston remains in this position.

From the beginning of the cycle the pressure p_3 has kept the blocking element closed. When the control valve closes ($t = t_5$) the water supply to the delay element is interrupted and the spring begins to move the piston DP upwards again, causing the water above the rolling diaphragm to drain to the outside through the restrictor R_3. After the control valve has closed the

pressure p_3 therefore falls back to the value βp_s and stays at this value until the piston reaches its highest position. During this whole cycle the blocking element remains closed and consequently the motor piston remains in its lowest position. The blocking element does not open until all the water has been expelled from the delay element and p_3 has therefore fallen to zero ($t = t_6$). The piston can then begin its upward stroke, forcing the water through the restrictor R_2. When the motor has again reached its uppermost position ($t = t_7$), the pressures p_1 and p_2 start to fall, until the supply pressure above the valve again becomes dominant and the cycle is repeated by the opening of the control valve.

The total duration of one cycle of the motor, i.e. the length of the interval between t_0 and t_7, depends on the speeds at which the pistons move. The speed of the motor piston during the downward stroke is determined by the pressure difference $p_1 - p_2$, the swept volume and the restrictor R_1; during the upward stroke it is determined by the pressure p_2, the swept volume and R_2. The speed of the delay piston depends on the pressure p_3, the volume swept by this piston and on the restrictor R_3. The design is such that the stroke of the delay piston lasts longer, so that the length of the cycle is mainly determined by the delay element.

The downward stroke of the motor piston is the working stroke of the motor, the linear movement of the piston rod then being converted by a ratchet-wheel and pawl into a rotation of the camshaft. The resistance to be overcome in this process is not the same for every working stroke, since during the rotation of the camshaft differing numbers of valves of input elements are opened, depending on the programme instructions. This varying load affects the speed of the piston and therefore causes a variation in the time taken for the downward stroke. However, to keep the cycle time as constant as possible, the downward stroke of the piston is made very fast to minimize the varying part of the cycle time.

In dealing with the logic scheme in fig. 7 we saw that the motor can be stopped by two diaphragm elements (*25* and *29*) after the motor, which interrupt the water flow through the motor on a signal from the thermostat. In fact these elements are not located after the motor, but between the blocking element and the restrictor R_2 (*fig. 11*). Here they work in the same way as the blocking element: as long as a high control pressure is applied to one of the elements, the flow remains blocked and the motor piston cannot begin its upward stroke. (We saw in fig. 7 that this high pressure arises when the appropriate input valve is opened, i.e. when the heating is switched on, and lasts until the desired temperature is reached.) The motor cannot resume

operation until all three elements have stopped blocking the flow. It would of course be possible to stop the motor by means of diaphragm elements in the supply line. In that case, however, the motor might stop during the working stroke at a moment when a valve was not properly opened or shut, which would upset the operation of the machine. In our design this is not possible, since the motor always completes its working stroke before stopping.
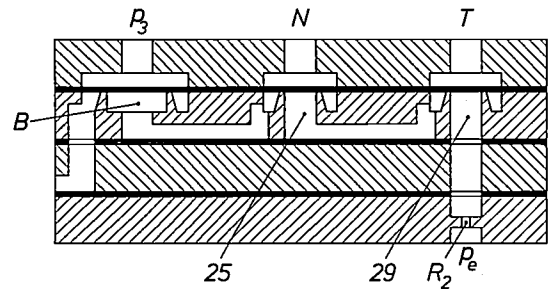
Fig. 11. Detail of the linear motor in fig. 9, showing that two additional diaphragm elements located between the blocking element B and the restrictor R_2 can stop the motor in the same way as the blocking element. These two elements (*25* and *29* in the logic scheme in fig. 7) are controlled by the output signals N and T from the thermostat. They block the working stroke of the motor while the water is being heated until the desired temperature is reached.

Pressostat and filler valve

Fig. 12 shows the design of the pressostat and the filler valve with the associated logic elements. The central part of the pressostat is the diaphragm *Me*, which is strengthened by a reinforcing disc. The space above this diaphragm is connected through the inlet *In* with the underside of the sump; the space below this communicates with the outside. After the machine has been filled, the diaphragm first closes the low-level valve *LL*, and then opens the normal-level valves *NL*. The water levels at which this takes place can be preset by special adjusting springs. The cross-section given in the figure is selected to show two of the three valves arranged in an equilateral triangle; these are *LL* in element *18*, with the adjusting spring *AL*, and *NL* in element *9*, with the adjusting spring *AN*. Only the top of the rod of valve *HL* can be seen. To operate the valves a fairly large force is required. Thus to close valve *LL*, it has to be forced down against the pressure *J* of about 1 atmosphere. Since valve *LL* must close when the pressure above the pressostat diaphragm is not yet greater than a few centimetres of water, the diameter of the diaphragm must be much greater than the diameter of valve *LL*.

The various elements and signals are indicated in fig. 12 by the same numbers and letters as used in fig. 7.

In the situation shown there is no water in the machine and input valve *III* and filler valve *FV* are closed. The pressure *D*, which sets this part of the control system in operation, is applied to the space above element *7* through a line not visible in the figure. The pressure from this space is applied to elements *18* and *19* of the

therefore acts as the input restrictor of baffle nozzle *11*.

It can be seen that the sandwich arrangement has been used throughout in the design of the parts of the control system described here, even for the large diaphragm of the pressostat and the diaphragm of the filler valve.
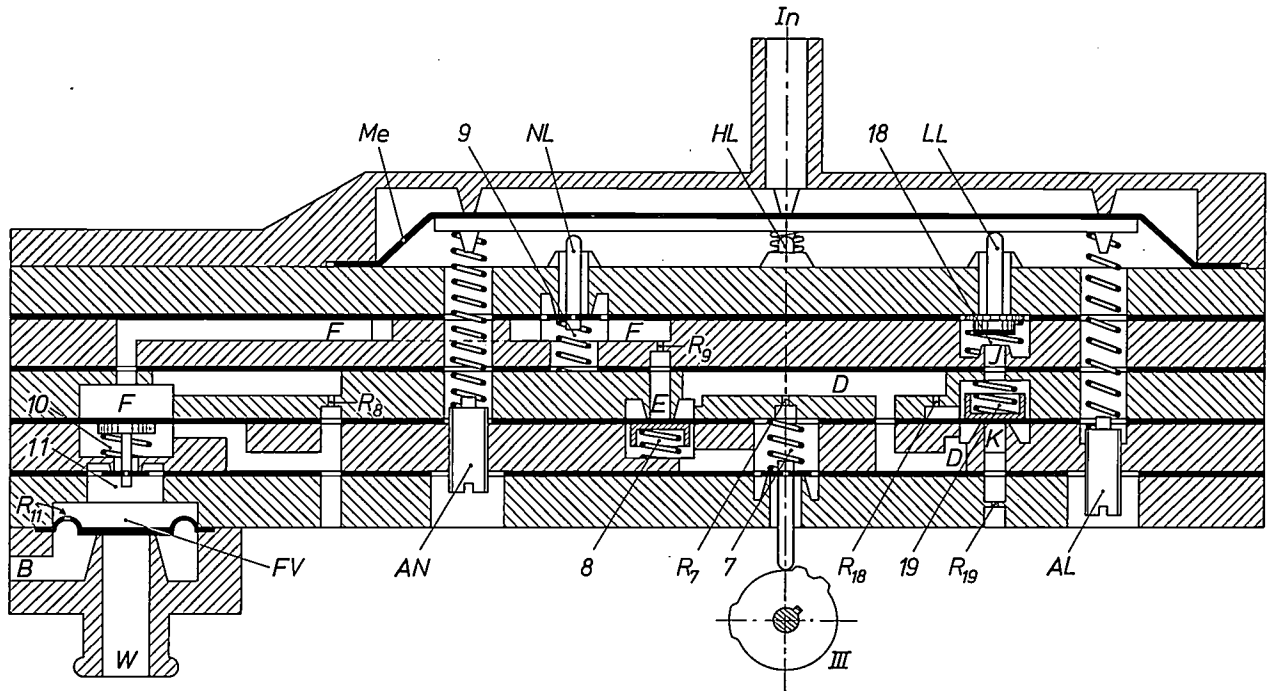
Fig. 12. Cross-section of the pressostat and filler valve with the associated logic elements. *Me* pressostat diaphragm. *In* inlet connecting the space above this diaphragm with the underside of the sump. *FV* filler valve, *B* water supply, *W* drainage to wash unit. *LL*, *NL* and *HL* are valves of baffle-nozzle elements that are opened or closed by the diaphragm when particular water levels are reached (low, normal and high). *AL* and *AN* are adjusting springs for setting the low and normal levels. (The elements for the high level are not in the plane of the drawing.) *III* input valve which receives the signal 'fill to normal level' from a cam. The various water-pressure signals, like the logic elements and the associated restrictors, are indicated by the same symbol as in the scheme in fig. 7.

low-pressure part of the pressostat, and to elements *7* and *8* near valve *III*. The output of element *8* connected with the baffle nozzle *9* in the normal-level part of the pressostat, and from here the signal *F* goes to elements *10* and *11* which actuate the filler valve. The input and output restrictors of all these elements are indicated in the figure; for reasons of design the output restrictor R_8 of diaphragm element *8* is not situated immediately beside this element. The filler valve *FV*, which opens and closes the passage to the wash unit (*W* = 1 or 0) when there is water pressure at the input (*B* = 1), is an enlarged version of the diaphragm element. The filler valve opens when the pressure above the diaphragm of this valve decreases, and this happens when the valve of baffle nozzle *11* is opened by the signal *F* = 1 applied via element *10*. The water feed to the space above the diaphragm of the filler valve goes through a small hole in this diaphragm, which

The thermostat

The thermostat follows the design most commonly used in washing machines. The measuring element consists of a sealed tube filled with a fluid with a high coefficient of expansion. Fixed to this element is a capillary tube in which the fluid moves in response to changes in the temperature of the element. This movement is converted by means of a tiny piston (an accurately fitting pin in the end of the tube) into a linear movement, which is used for successively opening the valves of baffle nozzles *23* and *28* (see fig. 7).

Design of the complete system

All logic elements, the pressostat and part of the motor are contained in a sandwich construction of six plates and five rubber sheets, together forming a block of 11.5 × 8 × 5 cm. *Fig. 13* shows a view of the block from underneath. The camshaft with the drive is fitted

to the sandwich construction at the lower left; the filler valve, with the water supply on the right and beneath it the connection for the wash unit, can be seen projecting from the block. The capillary of the thermostat is mounted at the left. The connections for the electrical switches are near the upper side of the

Control for an ordinary washing machine

A system of the type described here can of course also be used for controlling an ordinary washing machine. There is only one fundamental difference between such a machine and a dishwasher: the drum of the washing machine has to be set in motion at regular
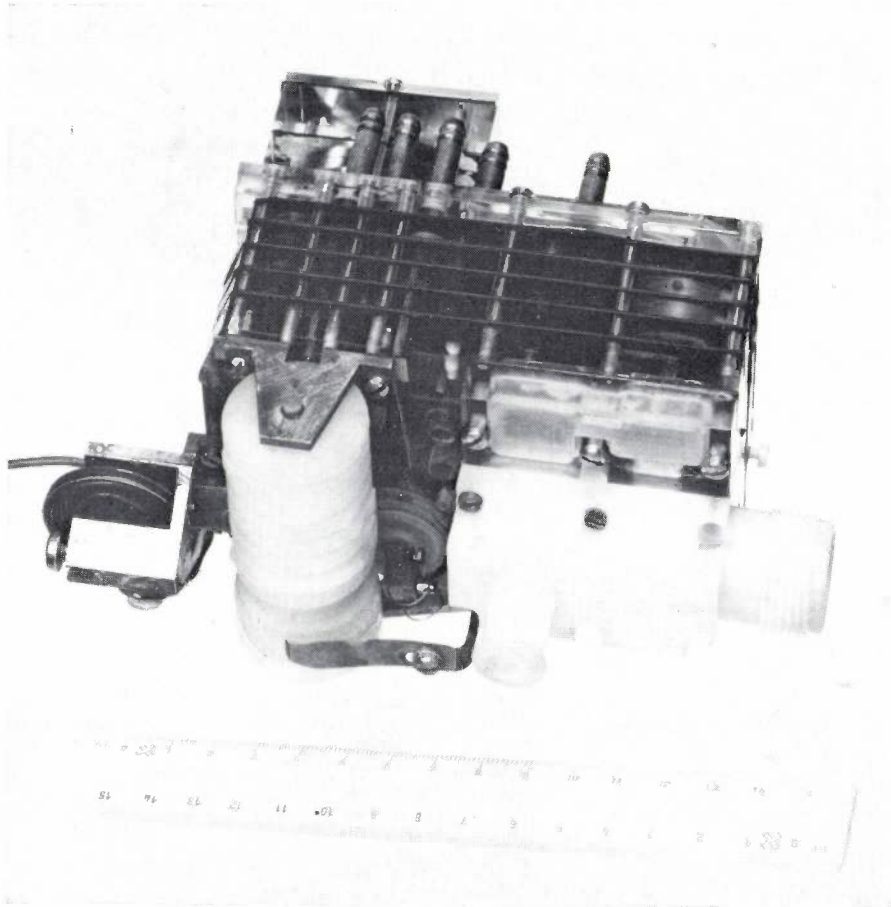
Fig. 13. View of the control system from underneath. The filler valve with the water supply and the outlet to the wash unit are on the right. To their left is the camshaft with the motor-driven ratchet mechanism. The connection to the thermostat is at the far left. The connections for the hydraulic output elements can be seen at the top of the system.

block on the left (not visible in the photograph) and the fluid lines for the door safety unit and for the supply of detergents and regenerating salt can be seen right at the top. The delay element is not located inside the block but projects above it.

Fig. 14 shows a view of the sandwich construction from above, partly dismantled. The valves, springs, etc. have been removed, while two plates and a rubber sheet have been folded outwards. The diaphragm seats of the baffle nozzles for the programme input can be seen in a row in the bottom plate. The large circular cavity beside it is the chamber for the motor piston. At the top, next to the block containing the delay element (see fig. 9), the pressostat diaphragm is visible through the top plate.

intervals, e.g. 12 seconds of clockwise rotation, 3 seconds stop, 12 seconds of anticlockwise rotation, 3 seconds stop, and so on. This washing motion requires a basic time of 3 seconds in addition to the basic time of 1 minute for the programme. This can be accomplished by using a linear motor with a cycle time of 3 seconds to drive a camshaft with cams to control the movements of the drum and an extra cam to give a signal after each twenty cycles. This signal (which thus appears once per minute) is used to actuate a second motor that rotates the programme timer through one step. Since the cycle time of the first motor is so short, the motor in fig. 9 may be used without a delay system; for the second motor all that is required is a piston with a spring and a rolling diaphragm.

The prototype described here has performed well in the laboratory. It was found to be important to use rubber of good quality for the diaphragms, otherwise the machine eventually develops faults as a result of permanent deformation of the diaphragms in the elements. Other problems might be expected from the restrictors which, with bores no more than 0.3 mm in diameter, could become blocked by scale formation or by solid particles or algae in the water. In our experi-

ments the water was fairly soft and of good quality, and this trouble was therefore not encountered. But even in less favourable conditions there is little fear of trouble from such blockages, since a water softener can always be added in hard-water districts and a filter can be used for trapping solid particles. In any case, the system can also be made to operate with static elements, which do not need restrictors. This approach is being investigated in our further studies.

Fig. 14. The sandwich construction of the control system; the plastic plates and rubber sheets have been stacked one above the other without the separate components of the logic elements (springs, valves, etc.). Two plates and a rubber sheet have been turned outwards to show details on the bottom plates such as the row of diaphragm seats for the baffle nozzles of the programme input, and the motor piston chamber. The block mounted on the top plate contains the delay element. The pressostat diaphragm is visible through the top plate.

Summary. The techniques of fluid logic, or fluidics, can be used in control systems for automatic washing machines. The systems operate from the water pressure and are simple and compact. The article describes a prototype designed for dishwasher control. After a brief survey of the various types of fluidic element (dynamic, quasi-static and static) a description is given of the elements used in the prototype, generally baffle nozzles and diaphragm elements. The elements are mainly of the quasi-static type, i.e. they use water in one of two states. Next a description is given of the operation of the dishwasher and of its programme of operations. In the control system the camshaft which operates the switches for the various parts of the programme is driven through a ratchet mechanism by a linear hydraulic motor with a cycle time of 1 minute. The mechanical input signals from the cams are converted by baffle nozzles into water-pressure signals which are handled in a fluid-logic system. The output signals of the control system are electrical (switches for pump, heating, etc.) and hydraulic (supply lines that provide the machine with water, detergents and regenerating salt). The practical basis of the design of the system is a sandwich construction of plastic plates and rubber sheets, in which the rubber sheets act as a seal and also form the diaphragms for the diaphragm elements. Such a system is very compact, and the whole control system measures only about $11.5 \times 8 \times 5$ cm. The system includes the linear motor, consisting of a piston with a valve-controlled rolling diaphragm, and the pressostat, consisting of a large diaphragm which, under the action of water pressure in the machine, operates baffle nozzles when specific levels are reached. The article concludes with an indication of the application of systems of this kind in ordinary washing machines.

The flow of highly rarefied gases

A. Venema

Some forty years ago Dr P. Clausing, who worked at Philips Research Laboratories from 1923 to 1960, published an article in 'Annalen der Physik' on the flow of highly rarefied gases. The appearance in 1971 of an English translation of Clausing's article renewed our interest in the subject, which remains of great importance to vacuum physicists. The article below by Dr Venema combines a modern treatment with a historical introduction which has special interest since Clausing's work is not always fully treated in all the textbooks.

Historical background

The year 1972 marked the tercentenary of the publication of Otto von Guericke's famous book '*Experimenta Nova (ut vocantur) Magdeburgica de Vacuo Spatio*' [1], in which he described his experiments with 'vacant space'. To remove the air from the space to be evacuated Von Guericke used piston pumps of a model directly comparable with the pumps that had been used for pumping water since ancient times. Originally he first filled up the space with water, and then pumped the water out. In this way he reduced the problems of removing the air to one of removing water, already a familiar technique. He also used a layer of water as a seal for pistons, cocks and couplings. In 1654 Von Guericke, who was then the burgomaster of Magdeburg, demonstrated the operation of his pumps with the well known Magdeburg hemispheres, which can still be seen in the 'Deutsches Museum' at Munich. Von Guericke may be assumed to have built the first vacuum pumps, and we may therefore look upon him as the father of vacuum technology, which thus goes back quite a long way in history.

An understanding of the flow of gases at low pressure was not obtained until the development of the kinetic theory of gases in about the middle of the last century. In 1909 M. Knudsen [2] published his experiments and calculations on this subject. He gave a formula for the flow of rarefied gases through tubes and openings for the case where the mean free path of the gas molecules (*fig. 1*) was equal to or greater than the radii of the tubes and orifices. Soon afterwards M. von Smoluchowski [3] published an improved version of the calculations, which gave the same results as Knudsen's,
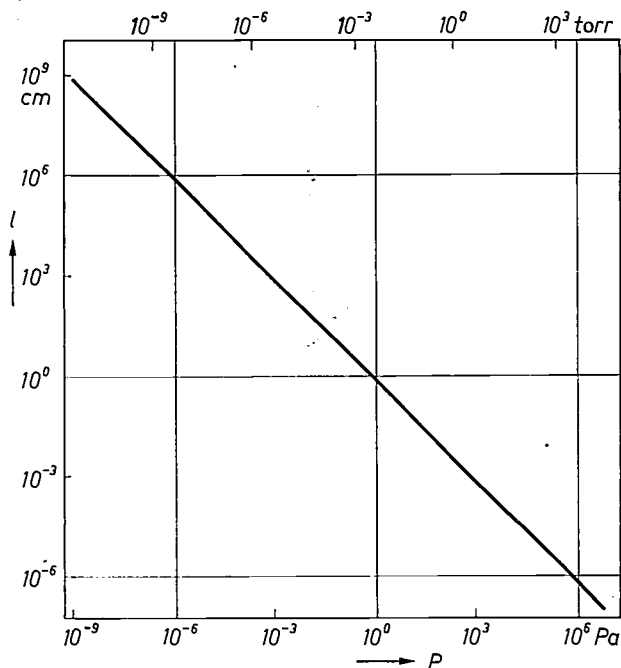
Fig. 1. The mean free path l in nitrogen as a function of pressure P (1 pascal = 7.5×10^{-3} torr).

however, when applied to cylindrical tubes. It had long been known [4] that gases under these conditions do not behave in accordance with the Poiseuille equation for viscous flow, but follow quite different laws. The reason for this is that as the gas becomes more rarefied

[1] Amstelodami, apud Joannem Jansonium à Waesberge, 1672. A German translation of this work, with the original illustrations, was published in 1968 by the VDI-Verlag, Düsseldorf.
[2] M. Knudsen, Ann. Physik (4) 28, 75 and 999, 1909.
[3] M. von Smoluchowski, Ann. Physik (4) 33, 1559, 1910.
[4] A. Kundt and E. Warburg, Pogg. Ann. 155, 337 and 525, 1875.

Dr A. Venema is with Philips Research Laboratories, Eindhoven.

the number of collisions between the gas molecules becomes considerably less than the number of collisions between the gas molecules and the wall. Eventually the gas becomes so rarefied that the molecules can no longer be said to be in collective motion. The behaviour of such a highly rarefied gas must be described on the basis of the individual behaviour of the separate molecules and no longer as the behaviour of a homogeneous viscous mass.

Following Knudsen, we speak in such a case of *molecular flow*. He gave the following expression for the molecular flow of a gas per second through a *long* cylindrical tube $(L \gg r)$ as a result of a pressure difference Δp:

$$Q_{pV} = \frac{4}{3} \sqrt{2\pi} \frac{1}{\sqrt{\rho}} \frac{r^3}{L} \Delta p, \qquad (1)$$

where Q_{pV} is the product of the pressure and the volume of gas flowing through per second, ρ is the density of the gas at unit pressure, r is the radius and L the length of the tube.

Knudsen gave two derivations for this formula. In one he calculated the tangential force exerted by the colliding molecules on the tube wall and compared this with the pressure difference across the tube, and in the other he calculated directly the number of molecules flowing through a cross-section of the tube. In both cases the tube was assumed to be so long that perturbations caused by the ends of the tube were negligible. It is therefore not surprising that for small values of L equation (1) does not approximate to the equation

$$Q_{pV} = \sqrt{\frac{\pi}{2}} \frac{1}{\sqrt{\rho}} r^2 \Delta p, \qquad (2)$$

which Knudsen derived for molecular flow through a circular opening of radius r in an infinitely thin wall.

Vacuum technology, which previously had led to no more than a number of remarkable demonstrations and interesting theoretical questions, began to acquire practical significance at about the turn of the century in the manufacture of incandescent lamps and X-ray tubes, and later also for thermionic valves. This prompted the development of more efficient vacuum pumps, based largely on the work of W. Gaede and I. Langmuir. In addition to mechanical pumps, in which the reciprocating pistons were replaced by continuously rotating components, diffusion pumps and vapour pumps were also developed, based on entirely new principles.

In the design of vacuum equipment for production purposes it was necessary to include in the calculations the pump-down time for a lamp or valve. This required an equation to describe molecular flow through a short

tube. In 1920 S. Dushman [5] published a formula that gave the relation between pressure difference and the gas flow per second for molecular flow through a *short* tube. Like Knudsen, he assumed that the expressions for a long tube and an opening could be written as:

$$Q_{pV} = \frac{\Delta p}{R \sqrt{\rho}},$$

where R is the molecular flow resistance. This resistance is determined entirely by the geometry of the tube or opening. Dushman now obtained his formula by postulating that the resistance R of a *short* tube ought to be the sum of the resistance of the entrance opening and the resistance of the tube as follows from Knudsen's formula for a long tube (eq. 1), giving:

$$Q_{pV} = \frac{\Delta p}{\left(\dfrac{3}{4} \dfrac{L}{\sqrt{2\pi}} \dfrac{L}{r^3} + \sqrt{\dfrac{2}{\pi}} \dfrac{1}{r^2}\right) \sqrt{\rho}}. \qquad (3)$$

This formula has the advantage of giving a satisfactory description both for large L and for $L = 0$. On the other hand, the use of Knudsen's formula (1) for the resistance of the tube completely overlooks the fact that it was derived for a long tube, whereas here it is used for a short one. Since Dushman gave no derivation for his equation, there was no way of estimating even roughly the reliability of the molecular-flow resistance calculated with his equation.

This unsatisfactory situation was resolved in 1932 when P. Clausing published an article [6] that gave an adequate treatment of molecular flow through a short tube. This forty-year-old article is still widely quoted in the literature. According to the Science Citation Index the references to this publication in the years 1968-71 averaged fifteen per year. To make this classic paper, originally published in German, more readily accessible to those interested in vacuum technology, an English translation was made available [7] through the courtesy of Veeco Instruments, Inc., one of the leading American firms in vacuum technology.

We shall now briefly consider Clausing's calculation of the molecular flow resistance of a short tube, and we shall then discuss the problems that arise in the calculation of this resistance for a number of short tubes or other components (valves, cold traps, etc.) connected in series. However, an understanding of molecular-flow phenomena is of more general importance to modern vacuum technology than simply as an aid to calculating the flow resistance of high-vacuum lines. We shall illustrate this in a brief discussion of the difficulties encountered in determining the pumping speed of high-vacuum pumps.

Molecular flow through a tube of arbitrary length

Gas-kinetic calculations

Clausing calculated the molecular flow in a tube connecting two vessels on the assumption that the vessels were so large compared with the transverse dimensions of the tube that the movement of the molecules in the vessels was not perceptibly disturbed by the entrance openings of the tube. This means that the cosine law for the angular distribution of the molecules applies to the entrance openings just as it does to the walls of the vessels (*fig. 2*). Since the movements of the molecules are completely independent of one another in molecular flow, the reasoning still remains valid when it is assumed for simplicity that the pressure in one of the two vessels is zero.
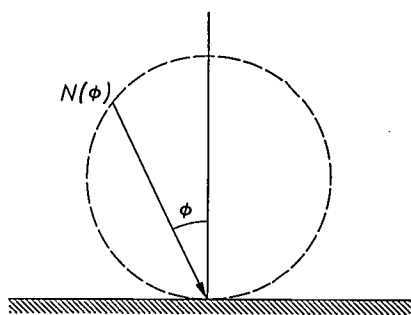
Fig. 2. Cosine distribution. The number of molecules, $N(\phi)$ striking a surface element per second at an angle of incidence ϕ with the normal to the wall, is given by $N(\phi) = N(0) \cos \phi$, where $N(0)$ is the number of perpendicularly incident molecules. The ends of the vectors that give the values of $N(\phi)$ as a function of ϕ lie on the periphery of a sphere of diameter $N(0)$.

Instead of the macroscopic quantities 'pressure difference' and 'mass of gas' Clausing used molecular quantities, i.e. the rate of incidence of the molecules and the number of molecules. He showed that all the Knudsen and Von Smoluchowski flow equations can be written in the form

$$K = WSv.$$

Here K is the number of molecules passing per second from the vessel at pressure p to the vessel at zero pressure, S is the cross-sectional area of the entrance opening of the tube, and v is the rate of incidence of the molecules (the number of gas molecules striking a surface per second per unit area) at this opening. He called the quantity W the *transmission probability*, i.e. the probability that a molecule entering the connecting tube will reach the second vessel without having returned to the first vessel. W is a dimensionless factor that depends solely on the geometry of the connecting tube; its value goes to 1 when the length

of the tube goes to zero. In calculating W Clausing assumed that the cosine distribution applied not only to the directions of incidence at the entrance opening of the tube, but also to the directions in which the molecules striking the tube wall leave the wall again. In other words, he assumed that the reflection of the molecules from the wall was completely diffuse.

It is not necessary to go into the physical details of the reflection of gas molecules from a wall to show that it is reasonable to assume a cosine distribution [8]. Clausing considered an arbitrary element of area dS inside a vessel which is uniformly filled with gas molecules and is in perfect temperature equilibrium with its surroundings. The molecules in this vessel move completely at random in all directions. The number of molecules that passes through an element dS at an angle θ to the normal will be proportional to cos θ because of the perspective shortening that applies to dS when seen from this direction. If now the element dS is replaced by a solid wall element, then the movement of the molecules in the environment of dS will only remain completely random if the molecules continue to leave dS in the same way as if there were no solid wall there. In other words, the molecules leave the wall in numbers that are proportional to cos θ. If this were not so, anisotropies and non-uniformities in the distribution of the gas molecules would occur within a layer equal in thickness to the mean free path of the gas molecules. With a suitable arrangement of the surfaces these inhomogeneities could set up a perpetual motion of the second kind, a 'machine' that performs work when operated from a single temperature and not from a temperature difference. Such a mechanism would contravene the second law of thermodynamics.

Although this reasoning shows that a cosine distribution of the molecules that leave the wall satisfies the second law, it does not follow that such a distribution is the only one possible. To prove this, however, requires a rather extensive treatment of the statistical mechanics of the scattering process at the wall [9].

To calculate W Clausing defined four probabilities:
1. The probability $w_{rr}(z)dz$ that a molecule leaving a ring of the tube wall will directly strike another ring of length dz at a distance z from the first ring.
2. The probability $w_{rs}(z)$ that a molecule leaving a ring of the tube wall will pass directly through a cross-section of the tube at a distance z.
3. The probability $w_{sr}(z)dz$, averaged over a cross-section of the tube, that a molecule leaving that section will directly strike a ring of the tube wall of length dz located at a distance z.
4. The probability $w_{ss}(z)$, averaged over a cross-section of the tube, that a molecule leaving this section will directly pass a second cross-section at a distance z.

Here 'directly' means without first colliding with

[5] S. Dushman, Gen. Electric Rev. **23**, 493, 1920.
[6] P. Clausing, Ann. Physik (5) **12**, 961, 1932, also Thesis, Leiden 1928.
[7] P. Clausing, J. Vacuum Sci. Technol. **8**, 636, 1971.
[8] P. Clausing, Ann. Physik (5) **4**, 533, 1930.
[9] E. P. Wenaas, J. chem. Phys. **54**, 376, 1971.

the wall again or with another molecule. The definitions are illustrated in *fig. 3*. It is not necessary to calculate all four probabilities separately. It can easily be shown that:

$$w_{rr}(z) = - \frac{dw_{rs}(z)}{dz},$$

$$w_{sr}(z) = - \frac{dw_{ss}(z)}{dz}$$

and

$$w_{rs}(z) = \tfrac{1}{2} r w_{sr}(z),$$

so that it is sufficient to calculate only one of the four probabilities.
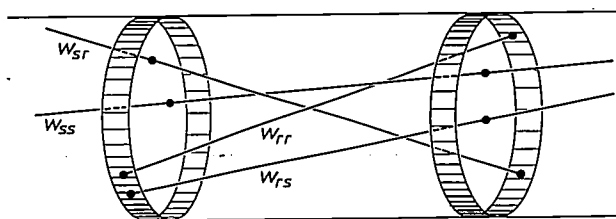
Fig. 3. Diagram illustrating the significance of the probabilities w_{rr}, w_{rs}, w_{sr} and w_{ss}.

Clausing calculated the probability $w_{ss}(z)$ for a cylindrical tube of constant diameter, using established calculations for the diffuse scattering of light, which of course also follows a cosine law. This led to an equation for the exit probability $e(z)$ for a molecule at a position z, with which he was able to calculate the desired transmission probability. Clausing gave a numerical approximation of the solution for this equation.

The results of the calculation are presented in *fig. 4* and also in *Table I*. They show that the theory for the transmission probability of a tube of zero length does indeed give the value for an opening in a thin wall. He also used the general equations to calculate the transmission probabilities for small rectangular openings in plates of finite thickness.

In 1956 W. C. DeMarcus [11] derived a transport equation for a Knudsen gas from Boltzmann's transport equation. This led him to an integral equation for the collision rate at any point of the wall; Clausing's equation for the exit probability proved to be a special case of this equation. By applying modern numerical methods to his integral equation DeMarcus was able to calculate the transmission probability for a short cylindrical tube. As can be seen from Table I, his results confirm those that Clausing had obtained 25 years earlier. The scale used in fig. 4 does not allow the results of Clausing and of DeMarcus to be presented separately.

Monte-Carlo calculations

It is now possible to use digital computers to calculate this probability statistically by the Monte-Carlo method. In this method the paths of the molecules travelling through the tube are calculated. A cosine distribution is assumed both for the directions in which the particles pass the entrance opening and for the directions in which they leave the wall after reflection.

The transmission probability is arrived at by calculating the paths of a very large number of molecules in this way. The more paths are calculated the more accurate are the values found. The vertical bars in fig. 4 give the regions in which there is a 95% certainty that the correct values of the transmission probability
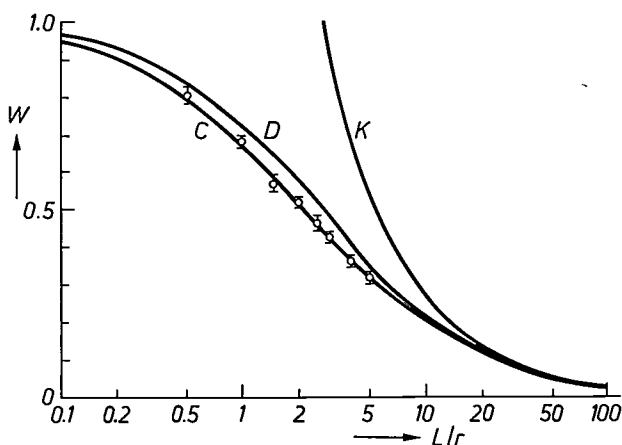
Fig. 4. The transmission probability W as a function of the length-to-radius ratio, L/r, of a cylindrical tube. C is the function $W(L/r)$ as calculated by Clausing, D is the same function calculated from Dushman's flow-resistance equation, and K is calculated from Knudsen's formula. The vertical bars show the intervals within which the transmission probability must lie with 95% certainty for a number of values of L/r from calculations performed by the Monte-Carlo method [10].

Table I. The transmission probability W for a cylindrical tube as a function of the ratio of the length L to the radius r of the tube, after Clausing [6] and DeMarcus [11].

L/r	W (Clausing)	W (DeMarcus)
0	1.0000	1.00000
0.1	0.9524	0.95240
0.2	0.9092	0.90922
0.3	0.8699	0.86993
0.4	0.8341	0.83408
0.5	0.8013	0.80127
1.0	0.6720	0.67198
1.5	0.5810	0.58148
2.0	0.5136	0.51423
3.0	0.4205	0.42006
4.0	0.3589	0.35658
5	0.3146	0.31053
10	0.1973	0.19099
20	0.1135	0.10938
50	0.0499	—
100	0.0258	—
1000	0.002658	—
∞	$8r/3L$	—

will lie. The length of the bars corresponds to four times the standard deviation σ.

Clausing's approach to the subject and his results are mentioned in many textbooks and publications. Some textbooks, however, still give Dushman's formula. Although this formula yields results that differ by less than 10-15% from the correct values and would thus be adequate for many applications, the nature of the formula, the explanation often given with it and the absence of any real proof do not contribute to a proper understanding of molecular-flow phenomena.

Several flow resistances in series

So far we have consistently referred to a single tube between two vessels. In practice, however, there are usually two or more sections of tube of different diameter, bends, cold traps or stopcocks connected in series. How can the resistance of such a combination of different components be found? Analogy with electrical resistance suggests that the resistance of the whole ought again to be the sum of the resistances of the parts. This is no longer true, however, when the smallest dimensions of the components are comparable with the mean free path of the gas molecules. To begin with, each component affects the flow in every other component. A molecule that has passed through a component will not go back into it again if there are no following components; if however there is another component then the molecule may go back into the first component. This is why the transmission probability of a group of components in series does not follow from those of the components. Another point is that the resistance of each component is calculated assuming a cosine distribution of the directions of the gas molecules at the input. In fact, however, molecules leave a tube in a beam whose width diminishes with increasing length of the tube. *Fig. 5* shows an example of this.

For the case of two tubes of identical dimensions in series *fig. 6* shows the actual value of the transmission probability W and the value calculated from $1/W = 1/W_1 + 1/W_2$, where W_1 and W_2 are the transmission probabilities of the individual tubes. The points were calculated from an approximation method due to C. W. Oatley [13].

Oatley has shown that the transmission probability for two consecutive tubes of *equal* diameter can be approximated by the expression

$$\frac{1}{W} = \frac{1}{W_1} + \frac{1}{W_2} - 1 .$$ (4)

[10] D. H. Davis, J. appl. Phys. **31**, 1169, 1960.
[11] W. C. DeMarcus, Union Carbide Nuclear Company, Oak Ridge Gaseous Diffusion Plant Report K 1302, Parts I-VI, Oak Ridge, Tennessee, 1956/57. For a short review of the results, see W. C. DeMarcus and E. H. Hopper, J. chem. Phys. **23**, 1344, 1955.
[12] P. Clausing, Z. Physik **66**, 471, 1930.
[13] C. W. Oatley, Brit. J. appl. Phys. **8**, 15, 1957.
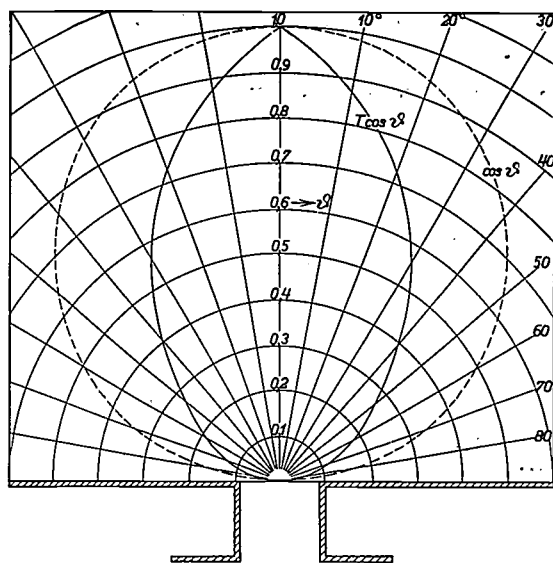
Fig. 5. Beaming of the molecules leaving a tube with $L/r = 2$ (solid curve). The dashed curve gives the cosine distribution [12]. The symbol ϑ in the figure, which has been reproduced directly from the article referred to in [12], corresponds to θ in the text.
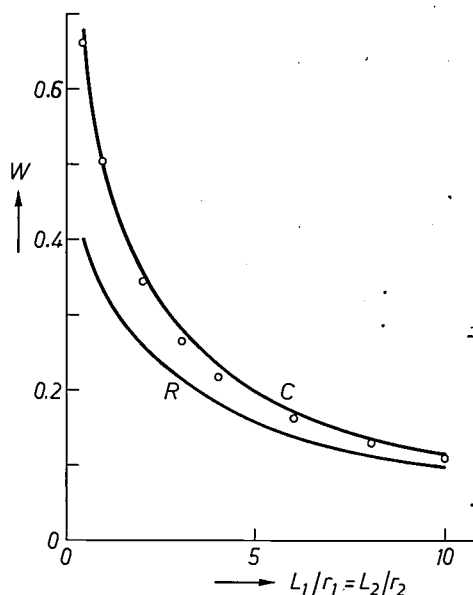
Fig. 6. Transmission probabilities W for two cylindrical tubes of the same length and diameter in series. C is the exact value following from the total length-to-diameter ratio, R the values calculated from $1/W = 1/W_1 + 1/W_2$ (analogous with electrical resistances). The points were calculated from Oatley's equation.

To derive this equation Oatley argued as follows. When N molecules enter tube I (*fig. 7*) on the left, a fraction W_1 will reach the right-hand end, while a fraction $1 - W_1$ will leave the tube again at the left-hand end. Of the NW_1 molecules that enter tube II a number NW_1W_2 will reach the exit, while $NW_1(1 - W_2)$ will again enter tube I. Of these molecules entering tube I, $NW_1(1 - W_2)(1 - W_1)$ will return into tube II, of which $NW_1(1 - W_2)(1 - W_1)W_2$ will eventually leave through the exit opening on the right. Examining further the fate of the remaining $NW_1(1 - W_2)(1 - W_1)(1 - W_2)$ molecules and summing all those that pass the exit opening of tube II, then

finally the following expression is obtained for the NW molecules passing through the two tubes in series:

$$NW = NW_1W_2 \{1 + (1 - W_1)(1 - W_2) +$$
$$+ (1 - W_1)^2(1 - W_2)^2 + \ldots\} = \frac{NW_1W_2}{W_1 + W_2 - W_1W_2}.$$

This gives the relation (4) between W, W_1 and W_2 immediately.

Equation (4) is an approximation, because it is assumed in the derivation that the directional distribution of the molecules that leave the first tube and enter the second one is a cosine distribution. This is not so; but, as argued by Oatley, the deviation from a cosine distribution is not large enough to give rise to serious errors. An indication of this is given in fig. 6.
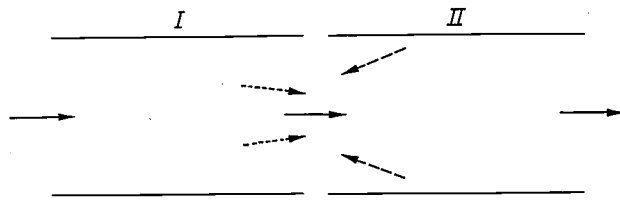
Fig. 7. In the derivation of equation (4), two tubes I and II of the same diameter are connected in series. The primary gas flow enters tube I from the left; successive 'generations' of scattered molecules are designated by dashed arrows.

An attempt to calculate the transmission probability for a number of arbitrary components in series leads to integral equations of the same kind as those already solved by Clausing for the simple case of a cylindrical tube. Solving equations of this type nearly always involves complicated numerical calculations. It is therefore often preferable to determine the transmission probability by the Monte-Carlo method [14]. However, useful approximate formulae for the transmission probability can be derived for relatively simple geometries [15].

Measuring the pumping speed of a high-vacuum pump

As Clausing pointed out, the concepts of transmission probability and hence of a resistance also apply to a vacuum pump. In many cases, however, and certainly in the case of a pump, it is more usual to use the reciprocal of the resistance concept, i.e. a conductance. To measure this conductance, i.e. the volume rate of flow of the pump, we must measure the quantity of gas flowing through per unit time (the throughput) and the pressure at the mouth of the pump.

The difficulty here is that in a flowing gas the pressure above the pump depends on the direction of measurement, so that the idea of a particular pressure at a particular place is meaningless without quoting a direction of measurement. Because of this the volume rate of flow of a pump can only be defined if the full

definition of the pressure at the pump mouth is given. To be able to compare the performance characteristics of different pumps it will therefore be necessary to reach agreement on the method of measuring the pressure. If a physically satisfactory choice is to be made, which links up as closely as possible with fundamental theory, a proper understanding of molecular-flow phenomena is essential. My suggestion at the time [16] was to connect the pump to such a large vessel that the directional distribution of the molecules at the pump mouth would be a cosine distribution, so that the pressure in the vessel could be taken as the pressure at the pump mouth. An arrangement of this kind, which gave good results for small pumps, is almost impossible for large pumps. Some of these have a mouth diameter of a metre or more, and the vessel, which is usually spherical, would have to have a diameter at least three times the pump mouth for reliable pressure measurements.

A way around this difficulty was found by using a vessel of the same diameter as the pump mouth and by performing the pressure measurement at a place in this vessel yielding numerically the same results as would
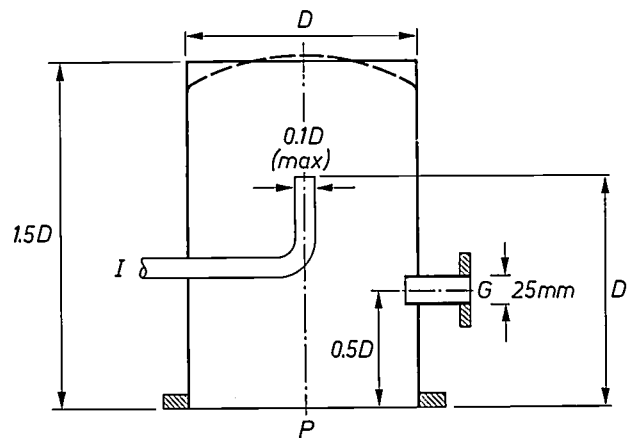
Fig. 8. Scale drawing of the vessel used for measuring the volume rate of flow (pumping speed) of a high-vacuum pump in accordance with ISO Recommendation R 1608. I inlet for gas flow. G pressure-gauge connection. At P the vessel is connected to the pump. Most dimensions are related to the diameter D of the pump mouth.

be obtained if a very large vessel were used. Monte-Carlo calculations by J. N. Chubb [17] showed that a system of this kind was indeed possible. Work by B. B. Dayton [18], H. G. Nöller [19] and W. Steckelmacher [20] also contributed towards the final choice of this measuring system. The last phase of the search for the most appropriate method for measuring the volume rate of flow of a pump took place at the meetings of the 'Technical Committee 112 — Vacuum Technology' of the 'International Organization for

Standardization' (ISO). The method recommended by this committee [21] for measuring the volume rate of flow of a high-vacuum pump is illustrated in *fig. 8*.

[14] See also L. L. Levenson, N. Milleron and D. H. Davis, Trans. 7th Nat. Symp. on Vacuum Technology, Cleveland 1960, p. 372, and J. N. Chubb, Proc. 4th Int. Vacuum Congress, Manchester 1968, Part 1, p. 433.
[15] Derivations of this type are to be found, for example, in B. B. Dayton, J. Vacuum Sci. Technol. 9, 243, 1972, and in L. Füstöss and G. Toth, *ibid.* 9, 1214, 1972.
[16] A. Venema, Vacuum 4, 272, 1954 (published in 1957).
[17] J. N. Chubb, Proc. 3rd Int. Vacuum Congress, Stuttgart 1965, Vol. 2, Part 1, p. 97.
[18] B. B. Dayton, Vacuum 15, 53, 1965.
[19] H. G. Nöller, Vacuum 13, 539, 1963.
[20] W. Steckelmacher, Vacuum 14, 103, 1964, and Vacuum 15, 249 and 503, 1965.
[21] ISO Recommendation R 1608, August 1970, Methods of measurement of the performance characteristics of vapour vacuum pumps, Part 1: Measurement of the volume rate of flow (Pumping speed).

Summary. A review of Clausing's contribution to the present understanding of flow phenomena at extremely low pressure (molecular flow) is preceded by an outline of the historical background. This includes a discussion of Knudsen's and Von Smoluchowski's formulae for the flow of highly rarefied gases through tubes sufficiently long that corrections for end effects are negligible. The industrial application of vacuum technology required formulae for shorter tubes. In 1920 Dushman proposed a formula for short tubes, but he gave no derivation or other theoretical basis. It was Clausing who provided a good theoretical description of molecular flow through tubes of arbitrary length. Later work by others in this field has fully confirmed the correctness of his approach.

The mean free path of gas molecules in molecular flow is large compared with the transverse dimensions of the tubes employed. Because of this it is not possible to determine the total flow resistance of separate parts of a line by simply adding the resistances of the components. In such cases an exact calculation can only be made by numerical methods. Approximate expressions are also available, however, that yield much smaller errors than simple addition.

Finally the importance of a proper understanding of molecular flow phenomena is demonstrated by the international agreement for measuring the volume rate of flow (pumping speed) of a high-vacuum pump.

A stabilized r.f. argon-plasma torch for emission spectroscopy

P. W. J. M. Boumans, F. J. de Boer and J. W. de Ruiter

Until recently the simultaneous multi-element analysis of aqueous solutions was none too easy a problem, especially when the concentrations involved were extremely small. A revival of interest in atomic emission spectroscopy, however, makes the situation look more promising. This development, which may surprise many chemists, is partly due to the use of the plasma torch, a 'flame' of extremely high temperature (10 000 K) that is not based on chemical combustion. An outstanding feature of the torch described here is the automatic stabilization of the power dissipated in the plasma.

Analytical atomic spectrometry and the plasma torch

Nowadays a great deal of chemical research on materials relies on physical methods of analysis referred to collectively as analytical atomic spectrometry [1]. The word 'atomic' indicates the limitation to the quantitative determination of free atoms, that is to say elements. These methods therefore provide very little information about the bonding of atoms in molecules or crystals. The radiation used is not nuclear but the radiation that results from processes in the electronic shells of the atoms and ions. In the type of analytical atomic spectroscopy considered in this article the radiation used is contained in the wavelength region extending from the ultraviolet to the near infrared (0.1-1 μm). The sample for analysis is atomized, which is most simply done by heating it in a vaporized state to a temperature at which the molecules dissociate. The analysis is made after calibration, using reference standards. To obtain accurate results there should be close correspondence between samples and standards, both in chemical composition and in physical properties, since the various

processes that lead to the emission or absorption of radiation are affected by the composition and properties of the samples. These effects are known as matrix effects or interelement effects. They can reveal their presence as a decrease in the temperature of the source of excitation, giving a marked reduction in the dissociation of the molecules in which a particular element occurs.

There are three basic methods of carrying out an atomic spectrometric analysis: the emission, absorption and fluorescence methods. The simplest radiation process, the resonance transition, is used in *fig. 1a* to demonstrate the physical differences between the three methods. The block diagrams in fig. 1b illustrate the arrangement of the related measuring instruments. The atomic fluorescence method will not be considered in this article, since it is still too recent for any proper judgement to be made of its practical value in chemical analysis.

The various emission techniques can be classified on the basis of the types of emission source, such as arc discharges, flames, spark discharges and the plasma torch, which is attracting increasing interest. Atomic

Dr P. W. J. M. Boumans and F. J. de Boer are with Philips Research Laboratories, Eindhoven. J. W. de Ruiter is with the Philips Centre for Manufacturing Techniques, Eindhoven.

absorption techniques may be subdivided according to the primary radiation source used and the absorption cell. The primary light source is often a gas-discharge tube with a hollow cathode, made of the element to be determined in the sample. The absorption cell may be a flame or a tubular furnace. Provided that several elements do not have to be determined simultaneously, the absorption method is usually preferable. This is evident from the great popularity of atomic absorption spectrometers for routine analytical applications in chemical laboratories.

In recent years, however, analysts have become more demanding, and a need has arisen for rapid and relatively simple techniques for simultaneous multi-element analysis, enabling perhaps ten elements to be determined at the same time. The analysis of samples in solution is preferred, mainly because it reduces problems associated with the inhomogeneity of samples and the preparation of standards, so that some of the matrix effects are removed. The trace analysis of surface water for environmental control, the trace analysis of soil extracts in agriculture, and a wide variety of analytical work in materials research, such as the analysis of thin films, all impose requirements that differ from those met by existing instruments. What is wanted is an

instrument that combines the advantages of the flame — ease of operation, stability and good reproducibility — with those of a d.c. carbon arc — low detection limit and simultaneous multi-element analysis.

To produce an instrument with this combination of features an attempt was first made to increase the temperature in the absorption cell of atomic absorption spectrometers. The idea was that the resultant higher degree of molecular dissociation would improve the detection limit. It was found, however, that the emission that is always to some extent present in such an absorption cell could become so high that it would be possible to use the cell as an emission source with characteristics somewhere between those of an ordinary flame (1900-2200 K) and those of the carbon arc (5000-7000 K). It was therefore decided, rather hesitantly at first, to return to the relatively neglected method of atomic emission, using a high-temperature flame (2800-3000 K) as emission source.

At about the same time, T. B. Reed and others were working on a new method of maintaining an ionized gas (a plasma) at atmospheric pressure [2]. The method is based on the principle of inductive heating, which is employed to induce high-frequency eddy currents in the plasma. The power thus supplied to the plasma sustains the ionization and gives the plasma a high energy content. Reed was able to demonstrate temperatures in the plasma up to 16 000 K, which is a great deal higher than the temperatures of about 3000 K in the high-temperature flames currently used for flame spectrometric analysis. Temperatures as high as this not only cause many more molecules to dissociate into atoms, they also vastly increase the number of atomic excitations. Consequently the light emission can be several orders of magnitude greater. The efforts made to use this plasma in emission spectroscopy have proved very successful [3] and have led to the development of a new emission source, called a 'plasma torch'. This device is encouraging more and more chemists to return to the atomic-emission method of analysis.

A plasma torch consists in the main of a gas-feed system and a plasma-tube assembly surrounded by an r.f. coil. Another essential part of the equipment, though to some extent separate from it, is the aerosol generator, or 'nebulizer'. For heating the plasma by r.f. induction we have developed a special r.f. generator
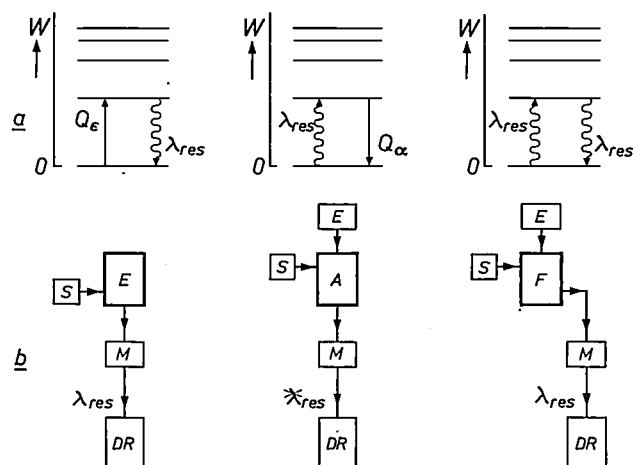
Fig. 1. a) The three radiation processes involved in the spectrometric methods of (from left to right) atomic emission, atomic absorption and atomic fluorescence, used for chemical analyses. W the energy of a free atom. Q_ε the amount of energy transferred to the atom on collision with another particle, resulting in the first excited state (the resonance transition). λ_{res} wavelength of the radiation which occurs on de-excitation — in the case of emission — or which gives rise to excitation — in absorption and fluorescence. Q_α the amount of energy which, on de-excitation, becomes available for conversion into mechanical energy (heat). b) Block diagrams of the spectrometric instruments based on these processes. M monochromator (usually a grating spectrometer). DR detector and recorder. S sample injection. E emission source. A absorption cell. F fluorescence cell. The fluorescence is observed at right angles to the path of the emitted light rays. The two other spectrometric arrangements are concerned with the presence (emission) or the attenuation (absorption) of radiation with the wavelength λ_{res}, in the direction of the emitted light.

[1] A review of analytical atomic spectrometry is given in: P. W. J. M. Boumans, Spektralanalysen, Techn. Rundschau No. 37, pp. 49-53, 3rd September 1971, and ibid., No. 43, pp. 33-37, 8th October 1971.
[2] T. B. Reed, Induction-coupled plasma torch, J. appl. Phys. 32, 821-824, 1961.
[3] G. W. Dickinson and V. A. Fassel, Emission spectrometric detection of the elements at the nanogram per millilitre level using induction-coupled plasma excitation, Anal. Chem. 41, 1021-1024, 1969.

designed to keep the power supplied to the plasma automatically constant. This constant power in particular has considerably widened the scope of the new torch for emission spectrometry. We have measured the detection limit for thirty elements, and have found very low values (10 to 5000 ng per gram of solid) if there is no interference from the matrix.

In the following we shall consider in turn the design and features of the new torch and its components. In doing so we shall touch on the physical background of the very large temperature difference between a flame and a plasma torch, and we shall briefly discuss the automatically stabilized r.f. generator. In conclusion we shall mention some experience gained with the new torch as an emission source in an experimental arrangement for atomic spectrometric analysis.

Design and operation of the plasma torch

Fig. 2 shows a schematic cross-section of the 'burner' in our plasma torch. The gas-feed system consists of three concentric quartz-glass tubes held in an acrylic glass base. The working gas is argon, which has the overriding advantage of not forming chemical compounds with the elements of a sample. An argon plasma is also very easily ignited and maintained. It is ignited by switching on a low-power auxiliary inductor while the argon is fed into the ignition tube (the second tube in the assembly). This generates a small 'pilot' plasma. The r.f. generator then takes over the power supply, causing the amount and size of the plasma to increase rapidly and the torch to come up to its working temperature. The temperature is then about 10 000 K inside the coil and about 6000 K just above it. The temperature decreases with increasing distance from the coil and is about 4000 K at a height of 30 mm above the coil. These temperatures, which are very much higher than in an ordinary flame, are easily reached because argon is a monatomic gas, which means that during heating no energy is lost to molecular dissociation.

The amount of energy needed to turn one mole of argon into a fully ionized plasma is 1.7 MJ. This means that in a torch in which argon at a pressure of 1 atm flows at the rate of 2 litres per minute, the argon will have to take up about 2.6 kW of energy to maintain the plasma. It can be seen from *fig. 3* that the temperature can then rise to a maximum of more than 16 000 K. For comparison the curve for nitrogen is shown; the maximum temperature for the same energy content is less than 11 000 K.

In an ordinary flame the temperature depends on the energy released on combustion. Above a particular temperature certain products of combustion start to dissociate, which uses up extra energy, so that the temperature rises no higher. The maximum temperature is therefore found in the range where strong molecular dissociation of the combustion products begins. The hottest flame (4850 K) is based on the reaction $C_2N_2 + O_2 \rightarrow 2 CO + N_2 + 1.082$ MJ/mol. The combustion products CO and N_2 therefore have the extremely high dissociation energies of 1.075 and 0.941 MJ/mol respectively, corresponding to 11.1 and 9.8 eV per molecule.

The temperatures in the plasma in the torch are so high that the plasma must not of course come into contact with a wall or other component. Here a significant advantage of inductive heating is that it requires no electrodes. The presence of electrodes would have a marked cooling effect and undesirable spectra could arise from evaporated electrode material.

In our torch the plasma is kept away from solid walls by means of a cylindrical sheath of a relatively
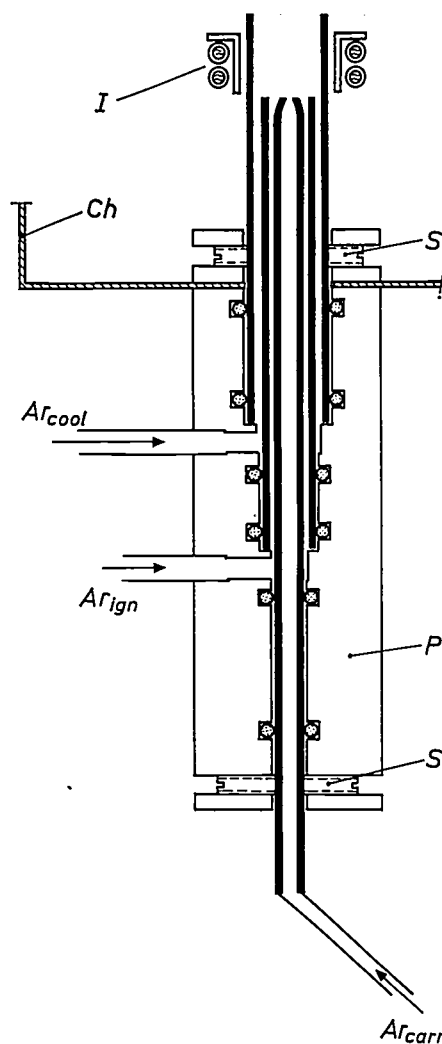
Fig. 2. The gas-feed system of the plasma torch. *P* acrylic glass holder in which three concentric quartz-glass tubes are mounted. The relatively cool shielding gas (Ar_{cool}) flows through the outer tube. The second tube is used for igniting the plasma (Ar_{ign}). The inner tube, the injection tube, carries the sample into the plasma as an aerosol in a carrier gas (Ar_{carr}). The inner diameters of the first two tubes are 18 and 12 mm. *I* watercooled r.f. coil. *S* screws for centring the tubes. O-ring seals around each tube make accurate centring possible and prevent leakage of the argon gas. *Ch* torch chamber with aluminium wall as an r.f. interference screen.
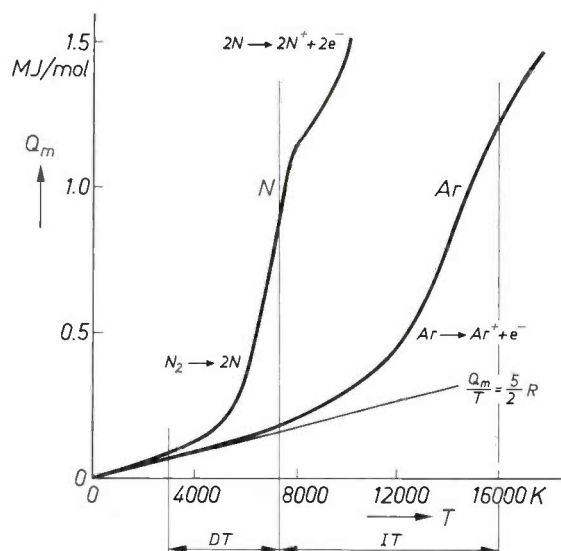
Fig. 3. The energy content per mole, Q_m, at 1 atm pressure for nitrogen and argon, as a function of absolute temperature T. DT denotes the region of dissociation temperatures, IT the region of ionization temperatures. The initial slope of the curve for argon is equal to $(5/2)R$, the specific heat at constant pressure of a monatomic ideal gas (R being the gas constant). For the same energy content the temperature of the argon is higher since there is no dissociation.

cool shielding gas, for which argon is again used because of its low thermal conductivity. The shielding gas is fed in through the outer tube at the rate of 15 litres per minute. Inside the tube the laminar flow of the shielding gas ensures that the plasma remains in position. In the region just above the tube the flow becomes turbulent, giving a sort of plume shape to the plasma. Between the coil and the outer tube a quartz-glass collar around the coil gives added protection against contact with the plasma.

The sample to be excited reaches the plasma through the injection tube (the inner tube in fig. 2). It is carried along in the form of particles in a stream of argon gas that flows through the injection tube during the operation of the torch. The flow rate of this gas (1 to 2 litres per minute) can be accurately controlled, which is necessary because it affects the performance of the torch as an emission source, as we shall see presently.

The frequency of the generator that supplies the power for maintaining the plasma is so high (about 50 MHz) that there is a distinct skin effect in the plasma [4]. Therefore, with a correct choice of the tube diameters we were able to produce a plasma that had its highest temperatures in a thick, roughly toroidal zone, whose exact shape is related to the flow rate of the gas. At the centre, where the particles of the sample reach the plasma, the temperature is much lower (*fig. 4*). The particles therefore have a much greater chance of penetrating into the hottest zone than they would if they were immediately confronted with it after leaving the injection tube [3].

Some idea of the thickness of this toroid can be obtained with the aid of an equation which states that the skin depth is proportional to $(\varrho/f)^{\frac{1}{2}}$, where f is the frequency of the r.f. field and ϱ is the resistivity of the plasma [5]. For an argon plasma with a temperature of 8000 K and a pressure of 1 atm the answer at 50 MHz is 0.3 cm, which is suitably small compared with the radius of the opening in which the plasma forms (0.9 cm).

To prevent the r.f. field from interfering with nearby electronic equipment, the plasma and the coil are housed inside an aluminium screen called the torch chamber. The housing of the r.f. generator and the torch chamber are rigidly interconnected, and are mounted together on a motor-driven platform that can be accurately adjusted to the desired height with respect to the fixed optical system. The optical part of our emission spectrometer is fixed; this greatly increases its reliability. A sliding flange provides a flexible connection between the torch chamber and the optics.

This arrangement makes it possible to determine simply and reproducibly the height in the plasma where the most suitable emission zone for an analysis is to be found.
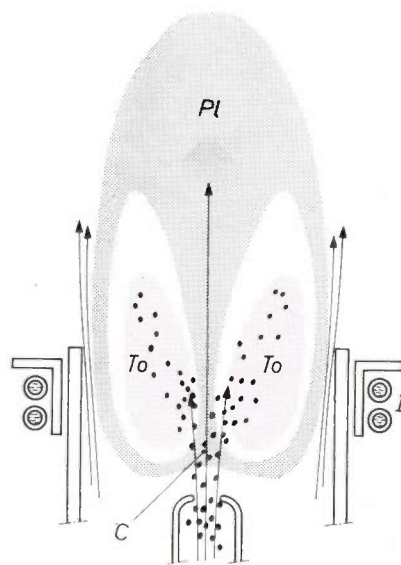
Fig. 4. Vertical section through the upper end of the gas-feed system, with plasma (*Pl*) and r.f. coil (*I*), which consists of two turns with internal water cooling. The toroidal structure of the hot core (*To*) of the plasma is due to the skin effect and to the carrier gas flow. The central zone (*C*) remains relatively cool. The particles to be excited are fed in through the injection tube; they can penetrate the hot region *To* relatively easily via the cool region *C*.

[4] The skin effect is treated in: H. B. G. Casimir and J. Ubbink, The skin effect, I. Introduction; the current distribution for various configurations, Philips tech. Rev. **28**, 271-283, 1967.
[5] For details see P. W. J. M. Boumans, Excitation of spectra, chapter 6 in: Analytical emission spectroscopy, Part II (ed. E. L. Grove), publ. Dekker, New York 1972, and also P. W. J. M. Boumans, Theory of spectrochemical excitation, Hilger & Watts London, and Plenum Press, New York, 1966.

The r.f. generator; power stabilization

Maintaining the plasma in our torch requires a power of no more than 1 or 2 kW. The quantity of power required was not therefore the main problem in the design of the r.f. generator. Far more important was the requirement that the power dissipated in the plasma should be very constant, otherwise fluctuations in the emission intensity could become very large. There are many possible causes of power fluctuations. In the first place of course there are heating-up effects in the generator itself, and the properties of the plasma may not always be constant. In particular the injection of the particles of the sample to be analysed may have a pronounced effect on the thermal conductivity of the plasma or on its resistivity, resulting in a marked change in the impedance. If these effects are not corrected in good time, they may even completely extinguish the torch. The r.f. generator [6] for our torch is designed to take the necessary corrective action automatically. We shall first give the general theoretical background of this stabilization, and then show in somewhat more detail how it is carried out.

The r.f. coil containing the plasma may be regarded as the primary coil of a kind of transformer. The plasma, which of course also has inductance, acts as the secondary winding; it constitutes a single turn, whose resistance is approximately equal to

$$R_{\mathrm{pl}} = \varrho \pi d/(h\delta).$$

In this equation ϱ is the resistivity of the plasma, which behaves approximately as a conducting cylindrical ring of diameter d, wall thickness δ (the skin depth) and wall height h. The coupling between the primary and secondary windings increases with the diameter of the plasma. The degree of coupling is determined by the factor k ($\leqslant 1$), which is equal to $M/(L_s L_{\mathrm{pl}})^{\frac{1}{2}}$, where M is the mutual inductance of coil and plasma, and L_s and L_{pl} are the inductances of the r.f. coil and of the plasma. Fluctuations in the energy content of the plasma affect the diameter d of the plasma through temperature changes; the situation resembles that of a gas at constant pressure and changing temperature. There is therefore a connection between the behaviour of the coupling factor and the power fluctuation. In general terms, the operation of the stabilization is that any small change in the power input, when a particular condition is satisfied, causes a new and compensatory change in the power through a change in the coupling factor.

Fig. 5 illustrates the operating principle of the r.f. generator under load, and also shows the operating point of the high-power amplifying valve used in the generator. The load consists of the internal resistance r_{p} of the generator and a resonant circuit which comprises

three components of the actual generator circuit — the capacitors C_1 and C_2 and the choke L — and also a choke L_t with a resistance R_t. Arranged in series, L_t and R_t represent the transformer constituted by the r.f. coil and the plasma. To understand the stabilization process it is necessary to know that during the growth of the plasma — which is accompanied by an increasing coupling factor k — the inductance L_t decreases.

Using the equivalent circuit for a loaded transformer we find:

$$L_t = L_s \left(1 - k^2 \frac{1}{1 + (R_{\mathrm{pl}}/\omega L_{\mathrm{pl}})^2}\right),$$

and

$$R_t \approx R_{\mathrm{pl}} k^2 L_s/L_{\mathrm{pl}},$$

where ω is the angular frequency of the r.f. current. In the derivation of the second equation the d.c. resistance of the r.f. coil is set equal to zero, and $\omega^2 L_{\mathrm{pl}}^2$ is substituted for $R_{\mathrm{pl}}^2 + \omega^2 L_{\mathrm{pl}}^2$. It can be seen from these equations that as k increases, the inductance L_t decreases while the resistance R_t rises.

The angular frequency, ω_c, of a Colpitts oscillator is given by:

$$\omega_c^2 = \frac{1}{(L + L_t)C} \left(1 + \frac{1}{1 + C_1/C_2} \frac{R_t}{r_{\mathrm{p}}}\right),$$

where $C = C_1 C_2/(C_1 + C_2)$. The resonant circuit itself has an angular frequency at parallel resonance (anti-resonance), ω_{ar}, that is a little lower than ω_c. Calculation shows that

$$\omega_{\mathrm{ar}}^2 \approx \frac{1}{(L + L_t)C} \left(1 - \frac{C_1 R_t^2}{L + L_t}\right).$$

Nevertheless, at the angular frequency ω_c, relating to the fundamental frequency in the pulses of anode current, the impedance of the parallel-resonant circuit is practically at its maximum. The higher harmonics also present dissipate very little power here, since the capacitance C_1 acts virtually as a short-circuit at these higher frequencies. The greatest transfer of power is obtained when the load is matched at the fundamental frequency.
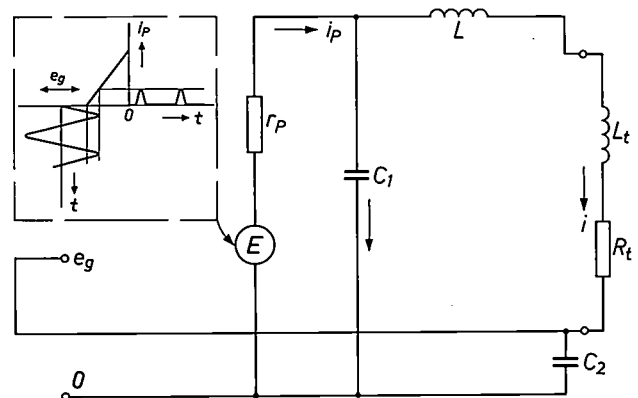
Fig. 5. Simplified equivalent circuit of the r.f. generator with load. r_{p} internal resistance of the a.c. voltage source E, describing the operation of a power triode. L_t, R_t equivalent circuit for the 'transformer', constituted by the r.f. coil and the plasma (as the secondary winding). C_1, C_2, L components of the generator. The generator is a Colpitts oscillator whose triode works as a class C power amplifier; the anode current i_{p} therefore appears as a series of pulses. The feedback voltage e_{g} across the capacitance C_2 sustains the oscillation. i current in the plasma.

An advantage of the circuit in fig. 5 is that the r.f. coil and the plasma form part of the resonant circuit in the oscillator. Whatever happens to the plasma, within fairly wide limits, the impedance of the resonant circuit at the fundamental frequency (ω_c) of the current i always remains at a maximum. (There can of course be a change in the fundamental frequency.) A separate resonant circuit for the coil and plasma, which is usually included in such circuits, was not necessary in our arrangement. Matching problems were thus avoided.

We shall now show how the variation of the coupling factor can give power stabilization. We shall see that for this purpose the magnitude of the inductance L must be such that L_t/L is smaller than 2. The Q (quality factor) of the loaded coil, i.e. $\omega_c L_t/R_t$, is given by:

$$Q \approx \frac{1}{R_t}\left(\frac{L}{C}\right)^{\frac{1}{2}}\frac{\lambda}{(1+\lambda)^{\frac{1}{2}}}.$$

In this expression λ represents the ratio L_t/L, and C the equivalent of the two capacitances C_1 and C_2 in series. The power stabilization is concerned with the power P_{pl} dissipated in the resistance R_t. Calculation shows that

$$P_{pl} \approx \alpha E^2 R_t C/L(1+\lambda).$$

In this expression α is a positive number smaller than 1 that gives a measure of the efficiency; E is the a.c. voltage of the generator, which remains practically constant. Using the factor $1/R_t$ we can substitute in the expression for Q the power P_{pl} dissipated in the plasma; this gives:

$$Q \approx \sqrt{\frac{C}{L}}\frac{E^2}{P_{pl}}\frac{\lambda}{(1+\lambda)^{3/2}}$$

Provided that α is known, this equation enables the combinations of Q and λ to be determined at which the plasma will take a power P_{pl}. In the Q,λ plane we can thus draw a family of characteristic curves with the power consumption as a parameter (*fig. 6*). As long as the operating point remains on the same curve, the power consumption does not change. The curves have a maximum at $\lambda = 2$; the value of the maximum is inversely proportional to P_{pl}. For any desired power at the same loaded Q of the coil there are apparently two values of λ, that is to say two possible operating points.

We shall now show that the situation is stable only at the operating point on the *left* of the maximum, i.e. at $\lambda < 2$. To do this we assume that at such an operating point the coupling factor k decreases slightly at a certain moment. The inductance L_t will then immediately increase slightly, and so too will λ. As a result the operating point shifts first horizontally towards a setting that gives the same Q at a higher power P_{pl}
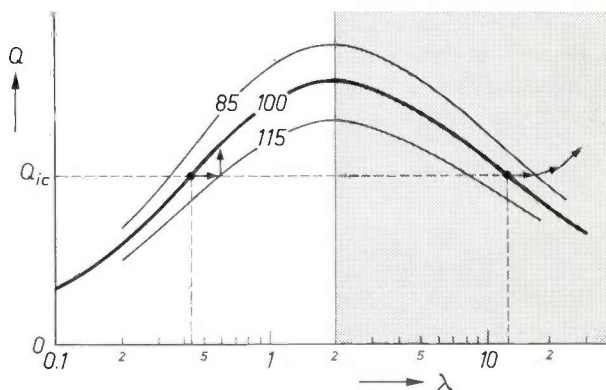
Fig. 6. The Q of the loaded r.f. coil as a function of the ratio L_t/L (see fig. 5), denoted here by λ. Each characteristic relates to one value of the power P_{pl} taken by the plasma (expressed in arbitrary units for the three curves shown). The a.c. voltage E of the generator is assumed to be constant. All characteristic curves have a maximum at $\lambda = 2$. For any desired power there are two settings for the same value of Q_{ic}, only one of which, with $\lambda < 2$, is stable. On a change in the coupling factor between r.f. coil and plasma the operating point therefore automatically returns to the original curve, so that the power consumption remains constant. At operating points on the right of the maximum this is not the case.

— which slows down the decrease of k. Next the operating point moves almost vertically upwards and returns practically to the original characteristic curve, at a somewhat higher value of Q. If, on the other hand, the operating point is on the *right* of the maximum, then upon a decrase of k the operating point will move to a position where P_{pl} is somewhat lower, *causing k to decrease further.* As a result the operating point moves further away from the curve on which it was originally situated; in other words, in this situation there is no power stabilization.

If k had originally increased slightly, then of course exactly the opposite would have happened in the two situations (since λ would decrease), but even then the operating point on the left of the maximum is the only stable one.

On ignition the operating point starts somewhere at the top right on a characteristic curve with very low power and high Q in the non-stable region. The operating point then moves rapidly to the stable part of the curve, corresponding to the final value of the power consumption.

Thus, by giving the inductance L of the choke in the generator a value higher than $\frac{1}{2}L_t$, we ensure that the power consumption remains constant during small changes in the impedance of the plasma, whatever they may be due to [7]. This is what makes our r.f. generator so outstandingly suitable for use in a plasma torch for spectrochemical emission analysis.

[6] Philips type 131202/01 HF table model with separate power-supply unit (output power variable up to about 2 kW).
[7] A similar situation was encountered in stabilizing the zone length in a zone-refining process for purifying semiconductors. See E. Buehler, Rev. sci. Instr. **28**, 453, 1957.

The nebulizer

In an analysis, as we saw earlier, the sample is usually introduced as an aqueous solution. The solution, whose salt content must be no more than a few tenths of a per cent, is nebulized to produce a fine distribution in the carrier gas, ensuring that the sample is fed uniformly into the plasma. An important requirement is that molecules of the solvent, which are present in considerable excess, must not enter the plasma. If this did happen, these molecules — in particular water molecules — would absorb so much energy that the limit of the stabilizing capacity of the generator would be exceeded, and it would then no longer be able to sustain the plasma. It is therefore necessary to dry the aerosol, in other words the sol must be stripped of the solvent before it reaches the plasma. This is done by first passing the wet aerosol through a heated pyrex-glass tube, which raises the temperature of the aerosol to about 200 °C, thus vaporizing the water and leaving the dissolved salts as solid particles. The carrier gas with the water vapour and the sample particles then flows through a water-cooled reflux condenser, which traps the water vapour by condensation. The dry aerosol emerging from the condenser finally enters the injection tube (*fig. 7*). Only about 10% of the aerosol is lost during transport through vaporizer and reflux condenser. The condenser is cooled with tap water.

The aerosol generator used is an ultrasonic nebulizer developed for other applications [8]; some relatively
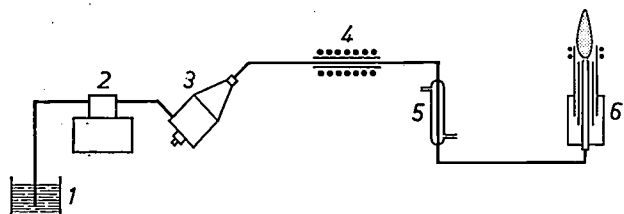
Fig. 7. Diagram of the plasma torch with nebulization system. *1* sample in aqueous solution. *2* pump. *3* ultrasonic nebulizer. *4* vaporizer. *5* cooled reflux condenser. *6* plasma torch. A good separation is obtained in the vaporizer at about 200 °C between the particles of salts and the solvent, which evaporates. The reflux condenser, cooled with tap water, traps the water vapour by condensation.

simple but essential modifications have been made to make the sample feed sufficiently fast and uniform. The nebulizing process has an efficiency of 25%, which means that 25% of the solution supplied enters the vaporizer. In the nebulizer the carrier gas flows tangentially into the space above the vibrator plate, and is thus brought into close contact with the nebulized solution. Complete mixing takes place in a conical fog chamber in the nebiluzer head, through which the wet sol finally passes to the vaporizer (fig. 7).

Experience with the new emission source

The performance of our plasma torch as an emission source was investigated in a conventional arrangement for spectrochemical analysis (*fig. 8*). Here we need only mention that the experimental equipment [9] includes two monochromators, which receive the emitted light alternately. The monochromators are each adjusted to a particular line of the spectrum, one emitted by an element to be analysed, and the other a reference line from an element added in known quantity. The intensity ratio of the two emission lines is used as the analytical signal. A dual-channel arrangement of two monochromators is a highly flexible spectrometer since it allows rapid and convenient adjustment of wavelengths. The resolution of these monochromators has to be high, since, owing to the extremely high temperature of the plasma, the emitted light contains a large number of lines, which have to be detected separately. A higher resolution also means better detection limits, since the intensity of the spectral lines is then increased in relation to the continuous background.

The investigations on the plasma torch have shown that the dissipated power remains constant within 0.5%. During the ignition of the plasma the resonant frequency shifts from the initial value of 51.8 MHz to 52.3 MHz at full load.

The temperature in the plasma depends closely on the distance of a particular zone from the top of the r.f. coil. This means, for example, that elements that are difficult to excite (and also have a high ionization potential) can best be detected in a zone of the plasma situated immediately above the coil, for the temperature is higher here than in regions further away from the coil.

Apart from the height of the plasma zone used, an important quantity is the flow rate of the carrier gas, which must be chosen with some care if the best detection limits are to be achieved. *Fig. 9* shows the optimum combinations of the location of the observation zone ('observation height') and flow rate for each of a number of elements.

It is usual to take the noise level and the slope of the analytical curve as a measure of the detection efficiency for a particular element. The analytical curve gives the relation between the concentration of the element in a series of standard samples and the signal from the spectrometer (*fig. 10*). In analyses with the plasma torch these curves are found to be straight lines at the small concentrations involved. When using an analytical curve it is necessary to bear matrix effects in mind that might introduce systematic errors.

The fluctuations in the background (the 'noise level') determine the smallest signal that can be said with a stated probability, say 95%, to be a 'true' signal, in
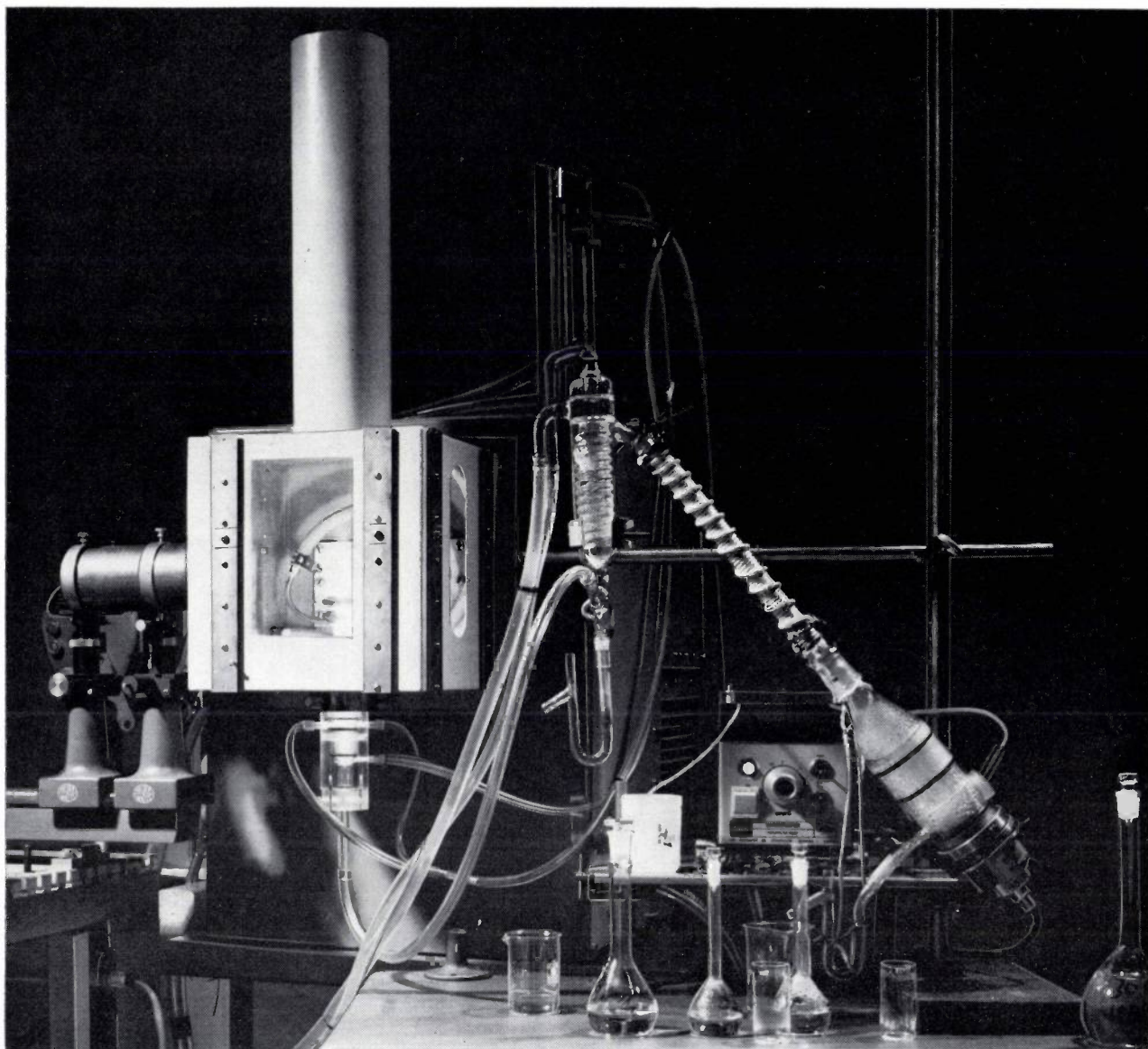
Fig. 8. Photograph showing general view of the plasma torch used as emission source. The aluminium torch chamber can be seen on the left. The 'flame' and the two horizontal turns of the r.f. coil are just visible inside the torch chamber. Part of the acrylic glass holder of the gas-feed system can be seen below the torch chamber. The upper (double) tube supplies the shielding gas, the tube below it the ignition gas, and the carrier gas with the particles for analysis flows in from below through the vertical tube. The nebulization equipment, including the ultrasonic nebulizer with fog chamber (right-hand corner), the evaporator and the reflux condenser can all be seen in the right foreground. The light emitted leaves the torch chamber on the left and passes through a lens system to the optical part of the spectrometer, which is not shown here.

other words a signal *not* due to background fluctuations. The noise level is found experimentally from a series of blank analyses, in which a solution is nebulized that does not contain the element to be detected but otherwise has as nearly as possible the same composition. The 95% signal corresponds to twice the standard deviation in the measured background. The slope of the analytical curve is determined from signals that are clearly above the noise level but are still of the order of magnitude of the background signal.

A good signal of course only gives the intensity of the line being analysed, which implies that the intensity of the spectral background always superimposed on analysis lines must be subtracted. It is the practice to work with a dimensionless measured quantity by dividing the intensity found by the intensity of a reference line. The advantage of such an intensity ratio is that

[8] LKB Medical AB, Ultrasonic Nebulizer NB 100 (frequency 3 MHz).

[9] See P. W. J. M. Boumans and F. J. de Boer, Studies of flame and plasma torch emission for simultaneous multi-element analysis, I. Preliminary investigations, Spectrochim. Acta **27B**, 391-414, 1972 (No. 9).

it is independent of interfering effects that affect both quantities to the same extent, such as a slow drift in the efficiency of the nebulizer. Moreover, by choosing the right combination of analysis and reference lines, it is possible to eliminate certain matrix effects.

A final point that we should make is that far fewer compromises have to be made in using the plasma torch in routine work then are needed for analysis with the d.c. carbon arc, even though the detection limits are
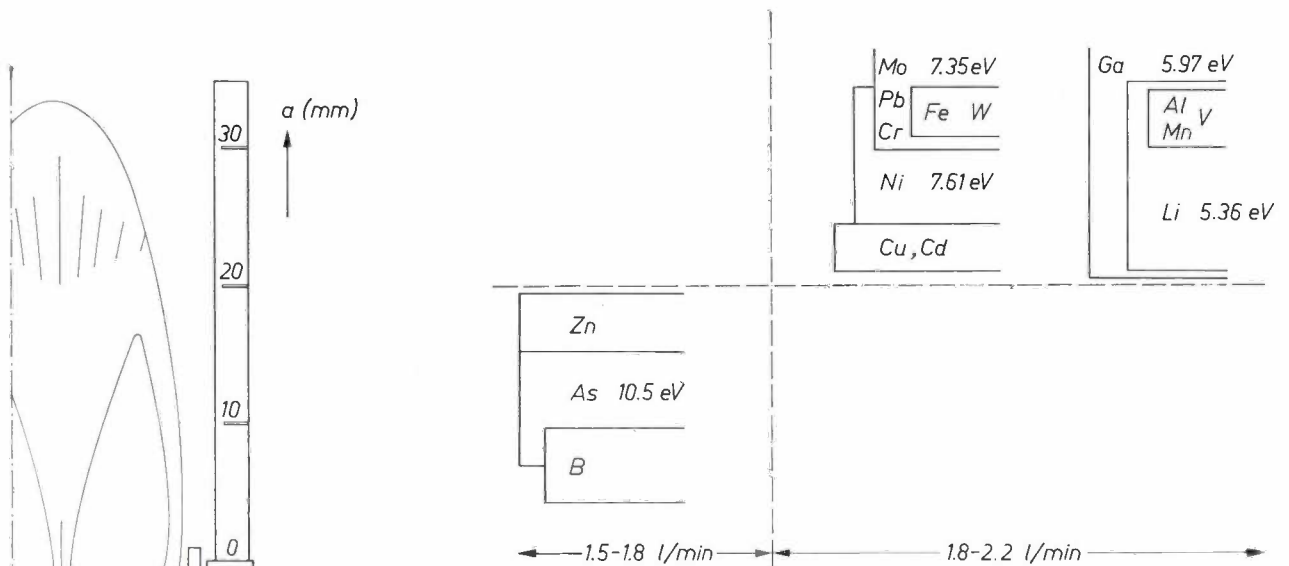
Fig. 9. The optimum combinations of carrier-gas flow rate and observation height a in the plasma for sixteen elements. The distribution of the elements in the diagram shows that simultaneous multi-element analysis is possible, provided the elements are separated into two groups. From the values of the ionization potential it can be seen that a high ionization potential is generally associated with a small value of a, i.e. with a high-temperature zone.

Table I presents our results for some thirty elements. The detection limit (C_{min}, fig. 10) was calculated in all cases as the product of the measured noise level and the measured slope of the analytical curve. Many of these detection limits are exceptionally low. It is reasonable to conclude that the combination of a stabilized plasma and our modified ultrasonic nebulizer with desolvation facilities has led to a marked improvement in the detection limit, in some cases by a factor of ten or more.

For multi-element analysis the optimum combination of the location of the observation zone in the plasma and the flow rate of the carrier gas must necessarily be a compromise between the requirements that apply for the various elements to be separately determined. The results obtained so far with our spectrometer indicate that a very good choice of observation zone and flow rate can be made by simply dividing the elements into two groups (see *Table II*, and also fig. 9). Conveniently it so happens that the detection limit for various elements does not change very much with the observation height in the plasma. This is because a slight deterioration of the signal amplitude then turns out to be compensated, and sometimes even more than compensated, by an improvement in the noise level.
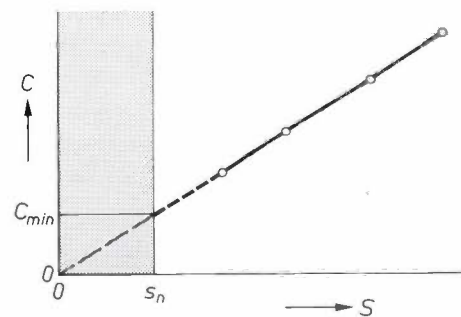
Fig. 10. Example of an analytical curve. The vertical scale shows the concentration C of the element to be detected, and the horizontal scale shows the signal S from the emission spectrometer. The points on the graph are obtained with the aid of a series of standard samples. s_n noise level, corresponding to the detection limit C_{min}.

just as low. The reproducibility of the results is also good (98 to 99% over a period of a few hours; the same values are found in analyses using a flame emission source).

To summarize briefly, we think that the most important features of our work are the improved stabilization of the torch and the modifications to the nebulizer, which make the equipment much easier to use.

Table I. Slope ($\Delta C/\Delta S$) of the analytical curve, and noise level (s_n) for thirty elements, measured with an emission spectrometer using the argon plasma torch as emission source. The detection limit C_{min} is set equal to the product $s_n \times \Delta C/\Delta S$ (see fig. 10). The signal S is dimensionless since the measured intensities are divided by a reference intensity (that of the background in this case). The noise level is also relative and therefore dimensionless. All tabulated values are for a measuring time of 15 seconds. There are four groups of elements: (1) elements for which the detection limit found is much better than the values reported in the literature, (2) elements for which the detection limit is improved by less than a factor of five, (3) elements for which, as far as known, no detection limit has been reported in the literature, (4) elements for which the detection limit is not so low. I spectral line originating from a neutral atom; II spectral line from a singly ionized atom.

Element	Wavelength of spectral line (Å)	C_{min} (ng/ml)	$\Delta C/\Delta S$ (ng/ml)	s_n
(1) Al	I 3961.5	0.2	4.0	0.056
Ba	II 4554.0	0.02	0.8	0.028
Cd	I 3261.1	3	80	0.042
Fe	I 3719.7	0.3	5.5	0.058
La	II 4086.7	0.4	8.3	0.046
Ni	I 3524.5	0.4	7.7	0.050
Pb	I 4057.8	2	62	0.038
Ti	II 3349.4	0.2	7.1	0.028
V	I 4379.2	0.2	4.0	0.054
Zr	II 3438.2	0.4	18	0.022
(2) Ce	II 4186.6	2	50	0.041
Cr	I 3578.7	0.3	4.3	0.058
P	I 2535.6	70	17×10^2	0.042
W	I 4008.7	1	33	0.033
Y	II 3710.3	0.06	1.8	0.033
(3) Be	I 2348.6	0.4	5.6	0.078
Ca	II 3933.7	0.02	0.5	0.033
Cu	I 3274.0	0.1	4	0.035
Ga	I 4172.1	0.6	11	0.054
Ge	I 2651.2	4	1.4×10^2	0.030
Hg	I 2536.5	1	33	0.038
Li	I 6707.8	0.3	9	0.032
Mg	II 2795.5	0.05	1.7	0.028
Mg	I 2852.1	0.2	6.3	0.031
Mn	I 4030.7	0.06	1.8	0.031
Mo	I 3798.2	0.2	45	0.036
Na	I 5889.9	0.3	20	0.017
Pd	I 3634.7	2	50	0.042
Sn	I 3034.1	30	500	0.060
Yb	II 3694.2	0.04	1.3	0.033
(4) As	I 2780.2	360	10×10^3	0.036
B	I 2497.7	80	20×10^2	0.039
Zn	I 3345.0	16	6.3×10^2	0.025

Table II. The optimum position (see fig. 9) of the observation emission zone for multi-element analysis. I atomic line, II ionic line.

a_{opt} (in mm)	Type of spectral line	Element
5-15	I	As, B, Be, Ge, Mg, P, Zn
15-25	I	Al, Be, Cr, Cu, Fe, Ga, Li, Mn, Mo, Na, Ni, Pb, Pd, V, W
	II	Ba, Ca, Ce, La, Mg, Ti, Y, Yb, Zr

Summary. Spectrochemical analysis of dilute solutions by means of plasma-torch emission is a technique of increasing importance. An argon plasma torch with stabilized r.f. heating is described. The plasma has a working temperature of about 10 000 K at atmospheric pressure. The gas-feed system consists of an acrylic glass holder with three concentric quartz-glass tubes, one for the shielding gas (which keeps the plasma in place), one for the ignition gas, and one for the carrier gas (flow rate of 1 to 2 l/min for sample injection and plasma maintenance).

The r.f. generator (50-70 MHz, 0-2 kW) is a Colpitts oscillator; the plasma impedance forms part of the parallel-resonant circuit. The power dissipated in the plasma is stabilized by exploiting the fact that the coupling between the r.f. coil and the plasma depends on this power. The solutions are ultrasonically nebulized and then stripped of water vapour; the sample to be analysed reaches the toroidal hot region of the plasma via the relatively cold centre.

The noise level, slope of the analytical curve and detection limit for some thirty elements in an aqueous solution were determined in a dual-channel spectrometric arrangement using the torch as emission source. Multi-element analysis with low detection limits (for most elements 0.02 to 10 ng/ml in a measuring time of 15 s) and a short-term reproducibility of 98 % make the system suitable for routine determinations of metals in solution (environmental studies, trace analysis, analysis of thin films).

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, Netherlands   *E*

Mullard Research Laboratories, Redhill (Surrey), England   *M*

Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes (Val-de-Marne), France   *L*

Philips Forschungslaboratorium Aachen GmbH, Weißhausstraße, 51 Aachen, Germany   *A*

Philips Forschungslaboratorium Hamburg GmbH, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany   *H*

MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium.   *B*

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, Netherlands.

P. T. Bolwijn & R. A. van Doorn: On the selective erasure of cathodochromic image displays.
J. Physics D **5**, 896-900, 1972 (No. 5).   *E*

P. W. J. M. Boumans: Studies of sputtering in a glow discharge for spectrochemical analysis.
Anal. Chem. **44**, 1219-1228, 1972 (No. 7).   *E*

J. P. Chané & L. Hollan: Preparation and characterization of doped silicon lithium.
Solar Cells, Proc. int. Coll., Toulouse 1970, pp. 319-331; 1971.   *L*

P. A. van Dalen: Propagation of transverse electroacoustic waves in a piezoelectric plate of symmetry C_{6v} or $C_{\infty v}$.
Philips Res. Repts. **27**, 323-339, 1972 (No. 4).   *E*

P. A. van Dalen & C. A. A. J. Greebe: Energy considerations on the propagation and the generation of acoustic surface waves in piezoelectric media.
Philips Res. Repts. **27**, 340-349, 1972 (No. 4).   *E*

M. Davio, A. Thayse & G. Bioul: Symbolic computation of Fourier transforms of Boolean functions.
Philips Res. Repts. **27**, 386-403, 1972 (No. 4).   *B*

J. M. Goethals: Méthodes de décodage par décisions majoritaires.
Rev. CETHEDEC **8**, No. 27, 37-45, 1971.   *B*

D. Gossel: Nonlinear circuit synthesis with controllable gyrators.
Proc. IEEE **60**, 906-907, 1972 (No. 7).   *H*

U. Gross & R. Kersten: Automatic crystal pulling with optical diameter control using a laser beam.
J. Crystal Growth **15**, 85-88, 1972 (No. 2).   *A*

J. Haisma, J. M. Robertson & U. Enz: Direct observation of light-induced Bloch wall pinning.
Solid State Comm. **10**, 1021-1024, 1972 (No. 11).   *E*

D. E. Lacklison, J. Chadwick & J. L. Page: Photomagnetic effect in ferric borate.
J. Physics D **5**, 810-821, 1972 (No. 4).   *M*

F. A. Lootsma: Penalty-function performance of several unconstrained-minimization techniques.
Philips Res. Repts. **27**, 358-385, 1972 (No. 4).   *E*

A. Mircea, A. Farrayre & B. Kramer: X-band GaAs diffused IMPATT diodes for high efficiency.
Proc. IEEE **59**, 1376-1377, 1971 (No. 9).   *L*

F. Möllers & R. Memming: Electrochemical studies of semiconducting SnO_2 electrodes.
Ber. Bunsen-Ges. phys. Chemie **76**, 469-475, 1972 (No. 6).   *H*

B. J. Mulder: Computation of the thickness and optical properties of thin layers from reflection and transmission spectra alone.
Philips Res. Repts. **27**, 315-322, 1972 (No. 4).   *E*

R. G. Pratt, G. Simpson & W. A. Crossley: Acoustic-surface-wave properties of $Bi_{12}GeO_{20}$.
Electronics Letters **8**, 127-128, 1972 (No. 5).   *M*

D. J. Schipper, C. Z. van Doorn & P. T. Bolwijn: Preparation of cathodochromic sodalites.
J. Amer. Ceramic Soc. **55**, 256-259, 1972 (No. 5).   *E*

M. J. Underhill: Delay-stabilised variable oscillator.
Electronics Letters **8**, 115-117, 1972 (No. 5).   *M*

C. H. F. Velzel: Inverse Fourier Spectroscopy, Part I: Theory.
Philips Res. Repts. **27**, 297-314, 1972 (No. 4).   *E*

W. L. Wanmaker, J. G. Verriet & J. W. ter Vrugt (Philips Lighting Division, Eindhoven): Manganese-activated luminescence in Cd_2PO_4F.
Philips Res. Repts. **27**, 350-357, 1972 (No. 4).

The Philips PW 1100 single-crystal diffractometer

J. Hornstra and H. Vossers

The automation of intensity measurements in the X-ray analysis of crystals is not only a matter of operational convenience. In fairly complicated crystal structures the number of intensities to be determined may run into thousands (perhaps hundreds of thousands in protein crystals), and there is virtually no alternative. The necessity for automation is also related to the accuracy desired. Higher accuracy than that given by the photographic recording method can be obtained by measuring the intensities of the reflections electronically, for example by using a scintillation counter. However, the crystal and detector then have to be positioned correctly for each reflection. Manual setting even for a thousand reflections is not really a practical possibility.

There are various automatic diffractometer systems available. One of them is the 'PAILRED' system, previously described in this journal. The present article describes a new system, the Philips PW 1100 single-crystal diffractometer, which not only carries out automatic intensity measurements but also carries out the rather laborious preliminary routines automatically by means of a built-in computer. Another novel feature of this system is that the orientation of the crystal is monitored and automatically redetermined during the measuring process if the change in orientation becomes too great.

In describing this system we shall place more emphasis on the many new facilities which the built-in computer offers the user than on the construction of the actual equipment. The development of the PW 1100 diffractometer was a joint project of the Philips Industrial Equipment Division and Philips Research Laboratories.

Introduction

In the usual methods of determining crystal structures it is necessary to know the intensity of the reflections that occur when a single crystal is irradiated by an X-ray beam. The number of reflections may be in the thousands for quite a simple structure, and can reach hundreds of thousands for complicated structures like those of protein crystals. In a previous article in this journal [1] a diffractometer system was described that measures the reflections automatically. The automatic facilities of this system, called 'PAILRED' (an acronym for Philips Automatic Indexing Linear Reciprocal-space Exploring Diffractometer) are limited to the main operations. The orientation of the crystal, in which an axis of the crystal lattice is aligned parallel to the main axis of the diffractometer, is not done automatically. The measurement of the reflections is partly automated. Measurements are

made in series, and for each new series of automatic measurements several angles and constants have to be set to new values.

In working with 'PAILRED' it was found that the alignment of the crystal and the setting for an automatic run were increasingly made on the basis of a fixed schedule, and that the occasional errors could usually be traced back to simple arithmetical mistakes. There was therefore a very good case for automating these parts of the procedure as well. 'PAILRED' is not however very easily adapted for this automation. Apart from the two angles of rotation that are set automatically, there are another four that have to be set by hand, whereas it is known that in principle a rotation about a total of three or four axes is sufficient to generate and measure all reflections. Moreover, the system of logic circuits in 'PAILRED' is not very suitable for considerable expansion.

Drs J. Hornstra is with Philips Research Laboratories, Eindhoven; Ir H. Vossers is with the Philips Industrial Equipment Division, Almelo.

[1] P. G. Cath and J. Ladell, Philips tech. Rev. **29**, 165, 1968.
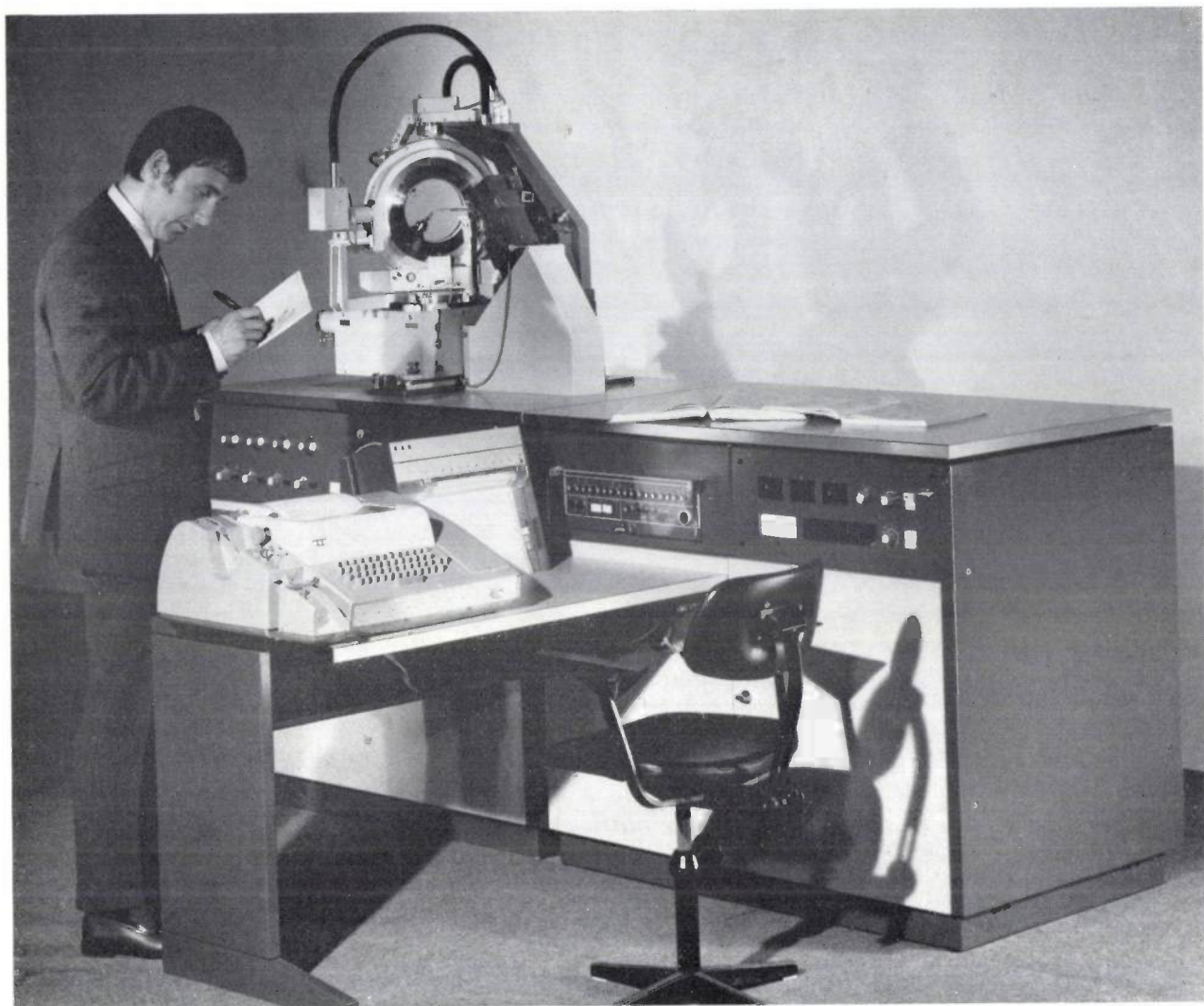
Fig. 1. The Philips PW 1100 automatic single-crystal diffractometer with built-in computer. The four-circle system in which the reflections are generated is located on top of the cabinet on the left. This cabinet contains the high-voltage generator. The X-ray tube is fitted in a holder on the right of the goniometer. The computer is located in the left-hand side of the second cabinet; the only part of it that can be seen is the control panel. The right-hand part of this cabinet contains the electronic circuits, together with a punch or a magnetic-tape recorder for the output. Some of the controls and the numerical display panels can be seen at the top with the count-rate meter, which displays the instantaneous value of the intensity of the radiation incident on the detector. There are two controls for selecting the quantities to be displayed on the numerical-indicator panels. A 'Teletype' and a pen recorder can be seen on the table in the foreground. The 'Teletype' is used both for input and output. It enters the parameters and starts the programs.

After inexpensive small computers for process control had appeared on the market it was clear that such a computer would be very suitable for controlling the new single-crystal diffractometer. On the initiative of the Philips Industrial Equipment Division, development work was started in 1967, resulting in the PW 1100 four-circle diffractometer, illustrated in *fig. 1*. With this system the preliminary routines just mentioned can easily be carried out in half an hour; with 'PAILRED' they require a day of fairly tedious work.

The equipment is arranged in two cabinets. The left-hand cabinet contains the high-voltage generator; the goniometer with crystal holder and X-ray tube are mounted on top of it. The left-hand part of the other cabinet contains the computer, of which only the control panel can be seen. The right-hand part contains the other electronic circuits and a paper-tape punch or magnetic-tape recorder for the output. This cabinet has numerical displays, which show various quantities that can be selected with the control on the right; these panels are continuously supplied with data from the computer. In the foreground there is a table with a 'Teletype' and a pen recorder. The 'Teletype' is used for both input and output. It is the main input device, used for starting all parts of the program and

for the input of all parameters. The control panel on the computer is used only for starting and stopping the equipment and for the program input.

There is a cooling system (not shown in the photograph) that cools the crystal under investigation and keeps it at low temperatures for days at a time by blowing it with boiling nitrogen vapour. To prevent ice from forming on the crystal holder, the stream of cold nitrogen is separated from the surrounding atmosphere by a 'sheath' of warm, dry nitrogen.

The design included a 'dedicated computer' right from the start. The computer was made an integral part of the equipment. This means that certain operations are assigned to the computer that could have been carried out by simple logic circuits. Arranging for them to be performed by the computer, however, gives greater flexibility. For example, a high speed of rotation about the four axes can be obtained with no loss of accuracy, since the computer ensures that the rotating parts switch to a lower speed just before the end of the rotation, so that the mechanical requirements are not too difficult.

With the computer it is unnecessary to give the crystal a particular orientation in advance. The lattice constants and orientation of an unknown crystal in an unknown orientation can be determined automatically in less than half an hour. Another interesting feature is that the orientation of the crystal can be monitored by the computer during the automatic sequence of measurements. As soon as the orientation changes by more than a preset limit value, it is re-determined within a few minutes, and the measurements are then automatically resumed.

The accuracy with which the PW 1100 measures the intensity of X-ray reflections was investigated on the spinel $ZnGa_2O_4$ [2]. Two small crystals were used, and large numbers of reflections were measured to determine the oxygen parameter. The standard deviation of this parameter was found to be 4×10^{-12} cm or 0.04 pm. For strong reflections the error in the intensity measurement was 1 to 1.5%; for weak reflections the error was determined by the counting statistics.

In the following we shall first deal with the four-circle geometry on which the diffractometer is based. We shall then briefly discuss the design of the equipment, and finally take a closer look at the various routines which the computer takes over from the user, i.e. the determination of the orientation, the measurement of the reflections and the accurate determination of the lattice constants.

[2] J. Hornstra and E. Keulen, Philips Res. Repts. 27, 76, 1972 (No. 1).

[3] See for example: U. W. Arndt and B. T. M. Willis, Single crystal diffractometry, University Press, Cambridge 1966.

The four-circle geometry

Principle

Since the principle of the four-circle diffractometer is now well known and has been described in detail elsewhere [3], we shall discuss it here only very briefly.

There are four possible rotations: three for the crystal and one for the radiation detector. *Fig. 2* gives a highly schematic picture of the mechanical part of our diffractometer. A vertical ring *A* is mounted between bearings at top and bottom. Inside this is a second ring *B*, which rotates in the plane of *A*. The mount *C* of the crystal holder *D* is fixed to the inside of *B*.

The four rotations are as follows:
1. The ring *A*, with everything it carries, can rotate about a vertical axis; this is the ω rotation.
2. The ring *B* can rotate inside *A*, the χ rotation.
3. The crystal holder *D* can rotate about its axis, the ϕ rotation.
4. The radiation detector *Det*, a scintillation counter as in 'PAILRED', can rotate about a vertical axis that coincides with the axis of rotation of the ring *A*: this is the 2θ rotation.
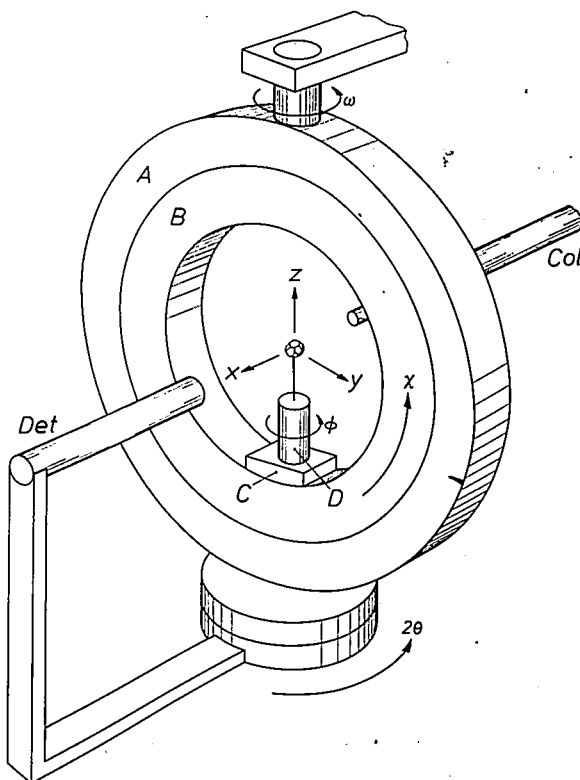
Fig. 2. Diagram of the four-circle system. Ring *A* rotates about the vertical ω-axis (ω rotation). Ring *B* rotates inside ring *A* (χ rotation). Ring *B* carries a mount *C* on which the crystal holder *D* is able to rotate about its long axis (ϕ rotation). The detector *Det* can rotate about an axis that coincides with the ω-axis (2θ rotation). The X-ray beam passes through the collimator *Col* perpendicular to the ω-axis and parallel to the *x*-axis. All parts drawn here are in the zero position.

As is customary, the angle θ is the angle between lattice plane [4] and incident beam for which the condition for reflection is satisfied (the Bragg angle). The angle 2θ is then the angle between incident and reflected beams, and this is the angle at which the detector must be aligned to receive the reflection — hence the name 2θ-axis for the axis about which the detector rotates.

Let us now assume that a crystal of known orientation and lattice constants is located at the centre of the instrument, at the point where the four axes of rotation and the incident beam intersect. We can then imagine that a particular lattice plane is brought into the reflection position in the following way (*fig. 3*). Starting from the zero position (see fig. 2) we first rotate the crystal about the ϕ-axis until the lattice plane perpendicular to CP_0 is parallel to the incident beam IB; we then rotate it about the χ-axis until the lattice plane is vertical. The incident beam is now still parallel to the plane, but from Bragg's law:

$$\lambda = 2d \sin \theta, \tag{1}$$

it must be at an angle θ (λ is the X-ray wavelength, d is the distance between successive lattice planes with the same orientation). If we now take ω equal to θ, the crystal will then be in the reflection position for the plane under consideration. When the detector is next rotated through the angle 2θ, the diffracted beam falls exactly at the centre of the detector aperture. Since the plane of the χ ring (the name we shall give to the assembly of rings A and B in fig. 2) now halves the angle between the incident and the diffracted beam, we shall refer to it as the bisector setting.

The four-circle geometry offers other possibilities. We have just discussed the setting at which $\omega = \theta$ in the reflection position. If the χ ring is prevented mechanically from taking up this position, it will be necessary to find another setting of ω at which reflection occurs. There are many such settings, all of them at different angles ϕ and χ. The essential difference between all these settings is that the crystal is rotated into different positions about the normal to the reflecting lattice plane. While the angle θ at which the X-rays meet the plane is always the same, so that the reflection condition is satisfied, the *azimuth* ψ varies. The way in which the angular settings ϕ and χ change when ψ changes may be understood as follows (*fig. 4*). When the crystal rotates about the normal to the reflecting lattice plane, the ϕ-axis must rotate with it. This axis describes a cone for which half the apex angle is equal to the angle between this normal and the ϕ-axis. The angles ϕ, χ and ω adjust themselves accordingly. The greatest deviations of $\omega = \theta$ occur when ψ is $\pm 90°$ and are equal to half the apex angle. One of these extreme positions is used in the automatic run when reflections have to be measured at $\theta > 70°$. This is because with these angles there is no room at the bisector setting for the χ ring between the X-ray tube and the detector. At this extreme setting, reflections up to an angle $\theta \approx 80°$ can be measured, provided the angle between the ϕ-axis and the normal is large enough.
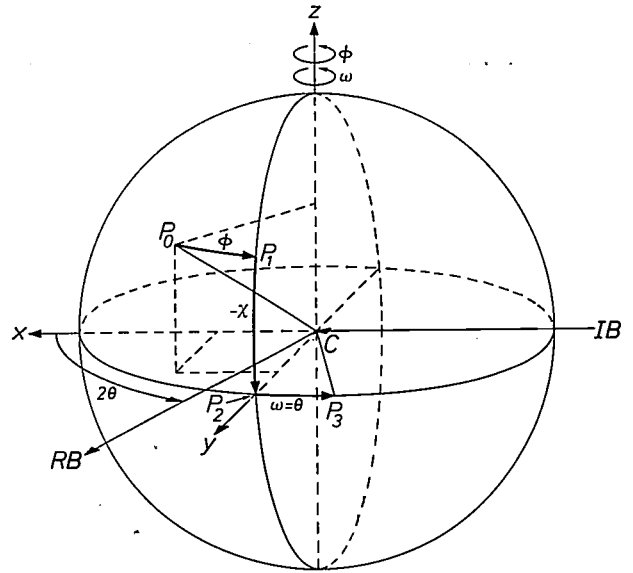
Fig. 3. Illustrating the setting for a reflection in the bisector position from the zero position of the instrument. In the sphere with centre-point C, the perpendicular from C to the plane that is to be rotated into the reflection position intersects the sphere at P_0. This point of intersection moves along the surface of the sphere to P_1 when the crystal is rotated about the ϕ-axis, from P_1 to P_2 when it is rotated about the χ-axis and from P_2 to P_3 when it is rotated about the ω-axis. In this position the incident beam IB makes an angle θ with the lattice plane being considered, and the beam is then reflected. The detector must be rotated through the angle 2θ to intercept the reflected beam RB. In this position the χ ring (the name given to the assembly of the A and B rings of fig. 2) bisects the angle 2θ, hence the name 'bisector position'.
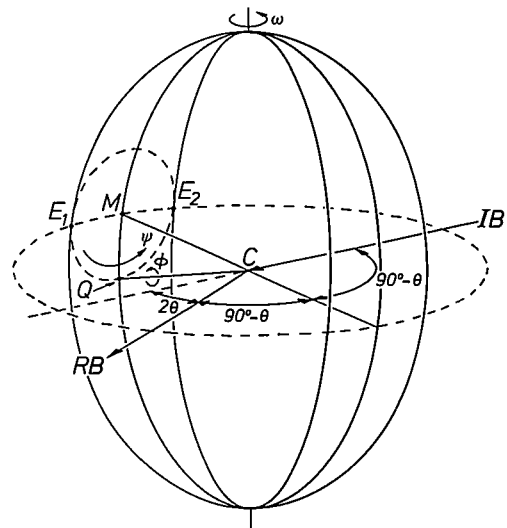
Fig. 4. Illustrating the azimuth rotation. The crystal rotates about the normal to the reflecting lattice plane. The angle between the incident beam and the lattice plane thus remains the same, and the condition for reflection is satisfied at all times. The crystal is located at the centre C where the incident beam IB cuts the ω-axis. MC is the direction of the normal about which the crystal and crystal holder rotate. The end-point Q of the ϕ-axis describes the dashed circle with the centre M. This point Q lies on the χ ring, which must therefore move with it. The ring is shown in three positions: the bisector position through M and Q, and the two extreme positions through E_1 and E_2.

There is in addition a special program for calculating the angles ϕ, χ and ω for any value of ψ and making the appropriate angular settings if mechanically possible. The use of this program will not be discussed further here.

Calculation of the bisector setting

To explain how the angles for the bisector setting are calculated, we need to use the concept of reciprocal space, which was introduced in the article of note [1]. We shall confine ourselves here to the essential features.

In reciprocal space each lattice plane is represented by a point at a distance λ/d from the origin in a direction perpendicular to the plane. The vector d^* from the origin to this point is the reciprocal-lattice vector corresponding to the plane. The points belonging to all possible lattice planes form a three-dimensional lattice, called the reciprocal lattice. This means that for each reciprocal-lattice vector we can write:

$$d^* = ha^* + kb^* + lc^*, \tag{2}$$

where h, k and l are integers, the 'Miller indices' of the lattice plane. The vectors a^*, b^* and c^* are the base vectors of the reciprocal lattice. The lengths of these vectors and the angles between them are characteristic of the crystal. They are related to the base vectors of the ordinary lattice as follows:

$$a \cdot a^* = b \cdot b^* = c \cdot c^* = \lambda;$$

$$a \cdot b^* = b \cdot c^* = c \cdot a^* = a^* \cdot b = b^* \cdot c = c^* \cdot a = 0.$$

To describe the orientation of the crystal in the goniometer we introduce a rectangular coordinate system that rotates with the crystal (fig. 2). If all the angles are zero, the x-axis points in the same direction as the incident beam, the z-axis is vertical, and the y-axis completes a right-handed orthogonal set of axes. We now resolve the three base vectors of the reciprocal lattice into their components along these axes, and assemble these nine quantities to form a matrix, which we call UB [5]:

$$UB = \begin{pmatrix} a_x^* & b_x^* & c_x^* \\ a_y^* & b_y^* & c_y^* \\ a_z^* & b_z^* & c_z^* \end{pmatrix}. \tag{3}$$

The coordinates of the point hkl in the reciprocal space or, in other words, the components of the vector d^*_{hkl}, are found from the following matrix operation:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = UB \cdot \begin{pmatrix} h \\ k \\ l \end{pmatrix}. \tag{4}$$

The relation between these coordinates and the angles for the bisector setting is now:

$$d^* = \sqrt{x^2 + y^2 + z^2} = 2 \sin \theta,$$
$$x = d^* \cos \chi \sin \phi,$$
$$y = d^* \cos \chi \cos \phi, \tag{5}$$
$$z = -d^* \sin \chi.$$

From this we can find ϕ and χ; we choose the solution where $-90° \leqslant \chi \leqslant 90°$. If the distance CP_0 in fig. 3 is made equal to d^*, we can then derive the equations (5) from the figure.

At this point we shall briefly describe the procedure carried out before starting the automatic intensity measurements. The crystal is centred by hand by translation in three orthogonal directions, i.e. it is moved into a position such that when the various axes are rotated no further displacement of the crystal is observed. With an arbitrary orientation of the crystal in the centred position we then operate the system to search for all the reflections, usually about twenty, in a small angular region. After the observed positions of the reflections have been converted to points in reciprocal space, the computer looks for the three shortest vectors between these points, not lying in one plane, and chooses these as base vectors of the reciprocal lattice. Once the reciprocal lattice is known, we can then set the instrument automatically for measuring the intensity of any desired reflection.

The equipment

In designing the mechanical part of the diffractometer a compromise had to be found between many requirements, some of them contradictory. In the first place we wanted high stability: the four axes were to intersect within a sphere of radius 5 μm. Secondly, we wanted highly accurate and reproducible angular settings; for the ω scan the accuracy was to be better than 0.005°, and better than 0.02° for the rotations about the other axes. The accuracy of rotation about the ω-axis had to be greater because we wanted to use this rotation for a highly accurate determination of the lattice constants (we shall return to this point at the end of the article). Thirdly, we wanted the angular region that could be scanned to be limited as little as possible by mechanical obstruction or by interception of the X-ray beam.

High stability is necessary to ensure that the crystal under investigation will always be irradiated by the same part of the X-ray beam. This is especially important when a crystal monochromator is used. (The PW 1100 works either with a crystal monochromator

[4] For convenience we have consistently used 'lattice plane' where we mean a set of parallel equidistant lattice planes.

[5] W. R. Busing and H. A. Levy, Acta cryst. 22, 457, 1967.

or with balanced filters.) If the monochromator is a perfect crystal, the intensity of the beam is fairly low; when a monochromator crystal with a large mosaic spread such as pyrolytic graphite is used, the intensity is greater.

The greatest stability problem was the method of mounting the χ ring to rotate about a vertical axis. This problem was solved by mounting the ring on bearings at top and bottom (see fig. 1). This arrangement gave the required accuracy without too much difficulty.

The rotation about the four axes is controlled by stepping motors ('Slo-Syn' HS 25) which make 200 steps per revolution. The motion is transmitted by gears, one step corresponding to a rotation of 0.005° about the ω-axis and 0.01° about the three other axes. A flexible coupling in the transmission ensures that the shocks from the motor are not transmitted to the mechanism, without affecting the accuracy even at high speeds. The maximum speed of rotation about the ω-axis is 12°/second with an accuracy of 0.002°, and about the other axes 25°/s with an accuracy of 0.01 to 0.02°.

The motor shafts are fitted with optical encoders, i.e. perforated discs situated between a number of lamps and photodiodes. These enable each step to be detected in the form of a pulse, and all pulses are fed to the computer. At every two hundredth step an extra pulse is generated. If an error has been made anywhere, this pulse does not occur exactly 200 pulses after the previous one. In such a case the measurement is interrupted and the shaft returns to the zero position. It then rotates again to the desired position. This closed-loop system prevents the accumulation of errors.

Special attention was paid to the rotation about the ϕ-axis. In view of the space available and the many electrical connections, we thought it undesirable to locate the motor that controls this rotation immediately below the crystal holder. The motor is situated outside the χ ring and its movement is transmitted by gears *in* the χ ring and a worm-gear. Another gear in the χ ring is connected to ring B and mount C and thus controls the angle χ. To achieve the desired accuracy it was also necessary to fit a counterweight opposite the base of the crystal holder. This arrangement enabled the χ ring to be kept narrow, giving little limitation in the range of angles that can be scanned; the ring subtends no more than 32° at the crystal. The goniometer head on which the crystal is mounted also comes within this angle (*fig. 5*).

Although the arrangements for adjustment of the X-ray tube have to meet the same requirements as in PAILRED, which also uses a crystal monochromator, adjustment is much simpler in the PW 1100. *Fig. 6* shows a vertical section through the mechanism. The
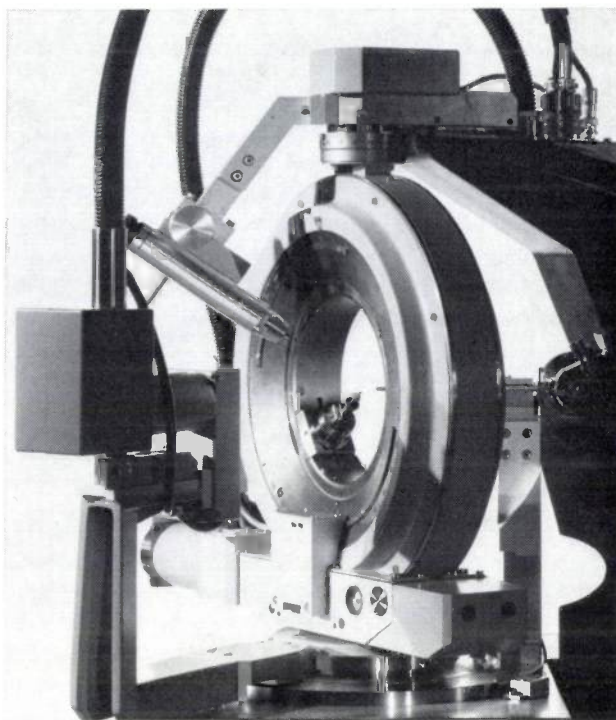
Fig. 5. The four-circle system. The detector can be seen on the left, with obliquely above it a telescope for aligning the crystal. The end of the collimator for the X-ray beam can just be seen in the opening of the ring.

tube holder T rotates about a shaft A_1 fixed to a plate Pl, which in turn rotates about a shaft A_2. The monochromator crystal Mo also rotates about the shaft A_2. Once this crystal has been set to the Bragg angle θ_{Mo}, the plate Pl is rotated until the beam IB passes through the centre point of the goniometer and strikes the
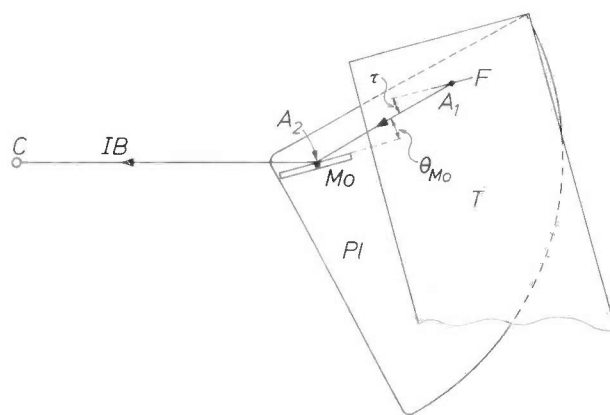
Fig. 6. Principle of suspension of the X-ray tube T when the crystal monochromator Mo is used. To produce a horizontal incident beam IB at the crystal C, the monochromator crystal must be oriented at an angle θ_{Mo} (the Bragg angle of the monochromator) and the axis A_1 of the X-ray tube must be brought into a line which makes an angle $2\theta_{Mo}$ with IB. The plate Pl can be rotated about A_2 to fulfil these conditions. The X-ray tube can be rotated about A_1. This rotation serves for setting the angle τ at which the X-radiation leaves the focus F.

crystal C. Rotation of the X-ray tube about A_1 varies the angle τ at which the useful X-ray radiation leaves the focus F.

Much of the electronics of the diffractometer consists of standard units that were not specially designed for the PW 1100. This applies in particular to the high-voltage generator (PW 1130) which only required a slight modification of the table surface to accommodate the heavy base plate of the goniometer, and to the computer. The computer storage capacity is 8192 words of 16 bits. Since control is left as far as possible to the computer, only simple input and output devices are required to the four stepping motors and to the optical encoders. The computer also controls a shutter, which stops the irradiation of the specimen when measurements are not being made. To determine whether the diffracted beam is exactly at the centre of the detector aperture, either the top half or the right-hand half of the aperture can be closed; the beam is central in the aperture if closing one half has the effect of halving the intensity of the beam. At high intensities attenuation filters are employed.

The intensity measurements can be carried out either with the aid of a crystal monochromator or with balanced filters. In this latter case the wavelength range in which the measurements are performed is determined by two filters, whose transmission is virtually identical over the whole range of wavelengths, with the exception of a small region in which the transmission is quite different. The filters are chosen in such a way that the desired $K\alpha$ line falls in the middle of this region. The measurement is performed first with the β filter, which absorbs very little of the $K\alpha$ radiation, and then with the α filter, which absorbs the $K\alpha$ radiation strongly. Subtraction gives the intensity that would be measured if only $K\alpha$ radiation were present. The computer can interchange these filters.

The counting system comprises a scintillation detector with a photomultiplier tube and a single-channel pulse-height analyser. The transmitted quanta are counted in a binary counter, which is read out by the computer every 1/32 of a second. An automatic correction is made in the counter for the 'dead' time.

Finally, measures were taken to prevent the moving parts of the goniometer from colliding with fixed parts of the equipment. The rotations about the ω-axis and the 2θ-axis, which could bring the χ ring or the detector into collision with the collimator, are monitored in the first place by the computer, which knows the limits of their range. Extra safety is provided by microswitches located beyond these limits. These stop the appropriate motor if they are touched. There is no risk of collision for the ϕ and χ movements, since the parts rotating about these axes are fairly small.

Software

We shall begin our discussion of the capabilities of the instrument by considering a number of operations that could easily have been carried out manually if this did not introduce some confusion for the computer. To avoid such confusion these simple operations also have to be performed by giving commands to the computer. Each command is started by typing a combination of three letters. If the computer knows the combination, it answers with two spaces. This may be followed by typing in a number of figures.

Fig. 7 gives some examples of simple operations, such as the stopping of a current command to the computer (STP), the interruption of a long-standing command (INT), the opening of the shutter (SHO), the setting of the α filter (SAF), the setting of ω, 2θ, χ and ϕ to particular values (OME, TTH, CHI and PHI respectively), and so on.

A somewhat more complicated command effects the automatic determination of the direction of the incident beam. This makes it unnecessary for the user to adjust the goniometer in such a way that the incident beam is exactly coincident with the χ-axis when $\omega = 0$. As soon as a reflection has been found, the command AZC (automatic zero correction) can be given. By measuring this reflection in various positions, the computer determines the direction of the incident beam.

STP				
INT				
SHO				
SAF				
OME	25			
TTH	50·5			
CHI	-13·6			
PHI	88·13			
TAN	25·000	50·50	-13·60	88·13
HKL	2 2 5 0			
PSI	10			
STM	60 =	8813		
WAV	0·7107			
XYZ	0·5 0·5 0·5 90			
TAN	-29·080	51·32	-90·00	135·00

Fig. 7. Example of some simple commands to the computer. After the command STP (stop) the computer immediately stops the command in progress; at the command INT (interrupt) it interrupts the program at an appropriate point. SHO is the command for 'shutter open'; SAF means 'set α filter'. The next four commands are for setting ω, 2θ, χ and ϕ at specific values. At the command TAN ('type angles') these four angles are printed out. If the UB matrix is stored in the memory, a particular reflection, in this case 225, can be set with the command HKL. The fourth figure indicates the azimuth; 0 indicates the bisector setting. The command PSI is used for changing the azimuth ψ of the adjusted reflection. The next command, STM (stationary measurement), enables a measurement to be carried out on the stationary crystal during a specific time (in our case 60 seconds). The equals sign precedes an output that will follow later, in this case the number of quanta counted. The wavelength of the radiation, here MoKα, is communicated to the computer after WAV. The command XYZ allows a setting to be made at an arbitrary point in reciprocal space; here again the azimuth must be stated. TAN is again used to obtain the corresponding values of ω, 2θ, χ and ϕ.

The deviations from the ideal direction are used as zero corrections.

The entire program for the PW 1100 is stored in the computer. This means that it is never necessary to change the program and all the facilities can be used at any time. A short and efficient program was obtained by writing it in the assembler language DAP-16.

Automatic determination of crystal orientation

Before automatic data collection can be started, it is necessary to know the lattice constants and the orientation of the crystal, contained in the *UB* matrix. The PW 1100 determines this from the unknown crystal in an arbitrary orientation. The computer performs the necessary operations so quickly that there is nothing to prevent this method from being used even when the lattice constants are known. We have already indicated the procedure: part of the reciprocal space is searched and all the reflections found are listed. After these reflections have been transformed into points in reciprocal space, the computer finds the three shortest vectors between these points, not lying in one plane, and takes these as the base vectors of the reciprocal lattice.

The region to be searched is determined by indicating a lower and an upper limit for θ, χ and ϕ. This is done after typing the command SPH (start peak hunting) at the same time giving the maximum number of reflections to be found, which can be up to 25 (see *fig. 8*). The region is searched along arcs of constant θ and χ, with alternately increasing and decreasing values of ϕ. At every reversal point χ is raised in value, and when it has reached its upper limit a repetition follows for a greater value of θ. The magnitudes of the steps in θ and χ depend on the horizontal and vertical dimensions of the detector aperture, which must be given in advance, following the command DET (detector) from fig. 8. The angle ϕ changes continuously at a rate that can be set by the operator.

The computer determines whether a reflection has been discovered by comparing the number of quanta counted with a discrimination level after every 0.25° of rotation about ϕ, an angular interval which is a little greater than the width of a reflection profile. A discrimination level is chosen which is a few times higher than the background level, but not so high as to run any appreciable risk of missing weak reflections.

After the command DIS (discrimination level) the selected rate and discrimination level are entered (see fig. 8).

As soon as the counting system has counted a number of quanta greater than the discrimination level in a particular angular interval of 0.25°, a few more steps are taken to find the maximum of the intensity distribution. Next, using the half-shutters that close the upper half or right-hand half of the detector aperture, a check is made to see whether the reflected beam passes through the centre of the detector aperture. If not, detector or crystal are rotated until the half-shutter exactly halves the intensity. Further peak location is then carried out in an omega scan and the aligning procedure is repeated in small steps. The angles θ, χ and ϕ ultimately found are printed out in a list with the peak intensity values. A check is made beforehand to see whether the reflection had been found earlier; if so, this is stated and the reflection is not listed a second time. The computer stores the reciprocal-space coordinates, calculated from equation (5).

The search procedure can be terminated in three ways:
1) when the stated number of reflections has been found;
2) when the whole angular region has been searched;
3) when the operator types in INT.

Next the computer calculates the reciprocal vectors between each point in the list and the origin, and between each pair of points. From these vectors the three shortest are chosen that do not lie in one plane. If a vector found later is almost identical with one of these three, the average is taken. Finally these three vectors are resolved into their components, printed out as columns of the *UB* matrix (fig. 8). Above each column the points are indicated between which the vector was first found and the number of vectors over which the average has been taken.

After the *UB* matrix the matrix $M = UB' \cdot UB$ is printed out. As a result of multiplication by the transposed matrix *UB'*, i.e. the matrix *UB* in which the columns and rows have been interchanged, the information on crystal orientation is lost, but the data on the lattice constants become much clearer. The matrix M therefore plays an important part in the determination of the crystal symmetry.

Each element of M is the scalar product of two base vectors in reciprocal space (e.g. $M_{12} = a_x*b_x* + a_y*b_y* + a_z*b_z*$). If two base vectors are equal in length, two diagonal elements are then identical. If two base vectors are at right angles, the corresponding matrix element is zero.

Owing to unavoidable errors of measurement, these relations are never satisfied exactly. (Although none of the elements in the example in fig. 8 is zero, the three base vectors are nevertheless found to be orthogonal.) The matrix M only gives an indication of the symmetry that could be present. Confirmation is necessary in all cases by comparing the intensities of reflections that are related by symmetry.

If the shortest base vectors are not parallel to the axes of symmetry, it is awkward finding the reflections that belong together. In such cases it is therefore the practice to introduce other axes that are parallel to the axes of symmetry. This makes the recip-

Fig. 8. Example of the output in the automatic determination of the crystal orientation. The initial input is the aperture of the detector, given after the command DET, and the required speed and discrimination level, given after DIS. The search for reflections is started with the command SPH (start peak hunting), followed by the maximum number of reflections to be searched for and the limits of the angular scan. The computer then types out the reflections found, θ, χ and ϕ, and the peak intensity. This is followed by the corresponding provisional UB matrix and the M matrix. After the command IRF (index and refine) the reflections are indexed and the matrix UB refined. The coordinates x, y and z are given for each reflection. After the command TLC (type lattice constants) the lattice constants are printed out.

```
DET   5 1·5
DIS   3 10
SPH   10 5 10   -30 30   0 60
  1      5·870      11·03      49·99      8130
  2      5·415      10·83      23·28      2883
  3      8·450     -28·47      21·35      2461
  4      7·940     -31·11      48·23      3656
  5      8·450      10·83      53·28       317
  6      7·650      11·21      35·26      1073
  7      6·535      10·69      18·66      4497
  8      7·715      10·38      15·41       723
  9      9·340     -23·58      35·64       542
 10      9·365     -24·24      18·14       326


   2   7   2   3   9   2   1   4   4

     -0·001678        0·079254        0·022491

      0·041473        0·001282        0·028259

     -0·006622       -0·009888        0·181915


      0·001767       -0·000014       -0·000070

     -0·000014        0·006381        0·000020

     -0·000070        0·000020        0·034398

IRF
  1   3   2   0    0·1537   0·1291  -0·0391    8130
  2   4   1   0    0·0732   0·1703  -0·0355    2883
  3   5   1   1    0·0941   0·2406   0·1401    2461
  4   3   2   1    0·1764   0·1575   0·1427    3656
  5   4   3   0    0·2314   0·1726  -0·0552     317
  6   5   2   0    0·1508   0·2133  -0·0518    1073
  7   5   1   0    0·0715   0·2119  -0·0422    4497
  8   6   1   0    0·0702   0·2546  -0·0484     723
  9   5   2   1    0·1733   0·2418   0·1298     542
 10   6   1   1    0·0924   0·2820   0·1336     326


     -0·001516        0·079174        0·022488

      0·042149        0·001354        0·028233

     -0·006409       -0·009912        0·181864


      0·001820        0·000001       -0·000010

      0·000001        0·006369        0·000016

     -0·000010        0·000016        0·034377

TLC      A =   16·660     B =    8·906     C =    3·833

      C0SA = -0·0011   C0SB =  0·0012   C0SC = -0·0002
```

rocal cell a few times smaller and the unit cell in the ordinary space lattice the same number of times larger. On the basis of this new (non-primitive) cell a reflection occurs only for particular combinations of h, k and l (e.g. in the case of a body-centred cell, only if $h + k + l$ is even). This is taken into account in the automatic series of measurements.

If the lattice constants for ordinary space are required, they can be obtained from the computer by giving it the command TLC (type lattice constants); see fig. 8 below. The computer then calculates them from the inverse of matrix M.

It may now be of interest to find the indices of the measured reflections based on the provisional matrix UB. These can be found by giving the command IRF (index and refine). From equation (4) it follows that the indices are obtained from

$$\begin{pmatrix} h \\ k \\ l \end{pmatrix} = UB^{-1} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

If a matching reciprocal cell is found, the indices will turn out to be almost integral. The computer rounds them off if the deviation is less than 0.125. Fractional indices may occur if too few reflections are found (the unit cell constructed then being too large), and also if the crystal is twinned or consists of several crystals.

The reflections whose indices are all integers are used for refining the *UB* matrix by the method of least squares. This is a linear problem if the coordinates x, y and z are treated as the observed quantities instead of the angles θ, χ and ϕ.

After refinement the accuracy of the *UB* matrix is usually sufficient for the whole automatic series of measurements. With the command OMT (omit) weak reflections that cannot be indexed and presumably originate from foreign crystal residues can be removed from the list, and strong reflections can be added to obtain a better distribution in reciprocal space by giving the command ADD. The same list is used for re-determining the crystal orientation during the measuring process.

The automatic measuring process

Before a start can be made on the automatic measurement of the reflections, a large number of parameters must be fed into the computer. Parameters already known to the computer, such as the *UB* matrix, do not have to be re-entered. It is desirable at this stage, by way of a check, to have a print-out of all parameters to be used. This can be obtained by the command PAR (parameters). *Fig. 9* shows such a print-out, and we shall use this list to discuss the various parameters.

Each reflection is scanned by means of a rotation about the ω-axis; the parameter SPE (speed) gives the speed of rotation in degrees per second about this axis. The width of the region scanned is given by SWD (scan width), followed by two quantities a and b, which occur in an equation giving the range $\Delta\omega$ of the scan: $\Delta\omega = a + b \tan \theta$. The region scanned is adapted in this way to the width of the reflection profile which increases with θ.

During the scan the detector may remain stationary, which is preferable for small values of θ, or can be made to rotate at twice the speed, which is preferable for large values of θ. The parameter SCT (scan transition angle) gives the value of θ at which a change should be made from one method to the other.

The insertion of absorption filters is determined by FLI (filter intensity level). If the peak intensity rises above the figure indicated with this command, the reflection is measured with suitable absorption filters.

In the case of weak reflections the relative error may become too great as a result of the counting statistics. To avoid this, the scan is repeated until the total number of counts exceeds the number specified with the command CNT (counts). In this way the weak reflections are measured more often than the strong ones, which improves accuracy and also saves time. The command MXN (maximum number) gives the maxi-

mum number of scans, to limit the time spent on measuring the very weak reflections.

Different methods of scanning are indicated by the command SMO (scan mode). The normal value of SMO is 1. When SMO is 0, only the peak intensity is measured; when SMO is 2, the measurement of weak reflections is stopped after the first scan, instead of being continued up to the maximum number of scans (MXN). The criterion here is whether the peak intensity I_{peak} in quanta per second is significantly above the background level I_{bck}, i.e. whether $I_{\text{peak}} - 2\sqrt{I_{\text{peak}}}$ is greater than I_{bck}. This saves time as well as limiting the quantity of output.

The manner in which the background on each side of the scanned region is to be measured is determined by BMO (background mode). If BMO is 0, measurements are made during a fixed period of time. If BMO is 1, measurements are made during half the scanning time. This is useful if the width of the scanned region or the number of scans is subject to variation. If BMO is 2, the background measurement time is adapted to the strength of the reflection. The stronger the reflection, the less the error in the background measurement contributes to the relative error in the net intensity, and the shorter the time spent on measuring the background. This also saves time for strong reflections.

The method of determining the background measurement time is based on the following. If I_s is the average intensity per second during scanning, and I_b the average intensity per second of the background, then the net intensity per second I_{net} is equal to $I_s - I_b$. If the scanning time is t_s and the background measurement time t_b, the variances are given by $\sigma^2(I_s) = I_s/t_s$, $\sigma^2(I_b) = I_b/t_b$ and $\sigma^2(I_{\text{net}}) = I_s/t_s + I_b/t_b$. It is known [6] that the minimum of $\sigma^2(I_{\text{net}})$ lies at $t_b = t_s\sqrt{I_b/I_s}$, subject to the subsidiary condition $t_s + t_b = \text{constant}$. This subsidiary condition does not apply in our case, since t_s is fixed. However, if in this case we look for the value of t_b at which $(t_s + t_b)\sigma^2 I_{\text{net}}$ is a minimum, or at which the inverse — the statistical weight per unit time — is a maximum, we find the same expression.

The commands for using the balanced filters are BAF = 0 or BAF = 90. The parameters TRG (theta range) are the lower and upper limits of θ. The parameter MVM (maximum value machine indices) will presently be discussed in connection with the matrix SEQ (sequence).

For crystals that have a high degree of symmetry it is not necessary to measure the whole of the reciprocal space; an octant or a part of an octant may be sufficient. This is indicated by the matrix BND (boundary), which when multiplied by the vector *hkl*:

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = BND \cdot \begin{pmatrix} h \\ k \\ l \end{pmatrix},$$

gives the three quantities p_1, p_2 and p_3. A reflection is

Fig. 9. Example of the output in an automatic series of measurements. The actual run begins after the command SDC (start data collection), followed by the initial values of the machine indices and the direction in which *H* has to be changed (by 1 or −1). First two reference reflections are measured (*RFA* and *RFB*). For each of these reflections the position of the maximum and the intensity are printed out. This line begins with the number of reflections measured so far, in this case none. Before the start, the command PAR produces all the parameters required for the automatic run; further particulars will be found in the text. In this example the lattice constants have also been printed out, after the command TLC.

```
PAR
    SPE  0·1000      SWD  1·40  0·00    SCT    45·00

    FLI   80000     CNT  50000    MXN   1

    SM0   1   BM0   1   BAF   90·00

    TRG    3·00    27·00    MVM  16   4   9

    BND   1   0   0
          0   1   0
          0   0   1

    SEQ   1   0   0
          0   1   0
          0   0   1

    ANT   0   0TP   1   0   1

    REF   2      60    0·010    0·050    10·00    0

              0   2   0  0·0300    0
             -3   0   2  0·0300    0

    SUB

      -0·001509      0·022329    -0·079107

       0·042072      0·028124    -0·001371

      -0·006187      0·181355     0·009993

    WAV  0·7107    DET  1·00   1·50    DIS  2·40    10

TLC     A =  16·702     B =    3·844    C =    8·912

        C0SA = -0·0005   C0SB =  0·0001  C0SC = -0·0035
SDC   1  0  0  1
      0 RFA  10·675      58060 RFB    5·835     42178
      3   0   0   0 0        34        93        42
      4   0   0   0 0        42       157        49
      5   0   0   0 0        39        76        24
      6   0   0   0 0        25      3934        23
      7   0   0   0 0        20        57        18
      8   0   0   0 0        16      2446        17
      9   0   0   0 0         5        22        14
     10   0   0   0 0        24      7119        20
     11   0   0   0 0        14        29        11
     12   0   0   0 0        10        60        14
     13   0   0   0 0         6        14         9
     14   0   0   0 0        11        29         8
     15   0   0   0 0         9        25         7
     16   0   0   0 0        23      5091        13
     16   1   0   0 0        26      4192        14
     15   1   0   0 0        10        12         6
     14   1   0   0 0         7        57         9
```

only measured if $p_1 \geqslant 0$, $p_2 \geqslant 0$ and $p_3 \geqslant 0$. In the example given in fig. 9, this means when $h \geqslant 0$, $k \geqslant 0$ and $l \geqslant 0$.

The matrix SEQ (sequence) is called by this name because it is used for determining the sequence in which the reflections are to be measured, i.e. which index must change first. To give the user this freedom of choice a second set of indices is introduced in addition to the normal indices h, k, l; these are the machine indices H, K, L. The relation between these sets of indices is given by:

$$\begin{pmatrix} h \\ k \\ l \end{pmatrix} = SEQ \cdot \begin{pmatrix} H \\ K \\ L \end{pmatrix}.$$

The sequence in which the machine indices change is fixed. First, H is changed by ΔH (which is $+1$ or -1) until $|H|$ reaches the maximum value, i.e. the first number after MVM. Then K is increased by 1 and the sign of ΔH is reversed (see fig. 9 below). When K has reached its maximum value, L is increased by 1 and K is reset to its minimum value. When L has reached its maximum value, the series of measurements is complete. Either the indices h, k and l or the machine indices H, K and L can be displayed on the numerical indicator panel.

[6] See the passage in small print on page 214 of the article by W. Parrish, Philips tech. Rev. **17**, 206, 1955/56.

Another important feature of the matrix SEQ is that it can be used for passing over absent reflections when a change has been made to a non-primitive lattice, for reasons that were explained on pages 68/69. The non-primitive lattice has a reciprocal unit cell that is too small. With a suitable choice of the matrix SEQ it is possible to ensure that the machine indices relate to a reciprocal cell of the correct size, i.e. to a primitive lattice. Since in any lattice there are many ways in which a primitive lattice can be chosen, the matrix SEQ provides considerable freedom of choice with regard to the sequence in which the reflections are to be measured.

If it is also desired to measure $\overline{hkl}$ immediately after each reflection hkl, we put ANT (anti-reflection) equal to 1, otherwise it is put equal to 0. These pairs have the same intensity when the structure possesses a centre of symmetry. If this is not the case, the absolute configuration can often be determined from the differences in intensity. The parameters OTP (output) determine the information required in a print-out and on the output tape.

A maximum of three reflections can be given as reference reflections, which are measured at fixed time intervals. In addition to their intensity, the position of the maximum of the profile is printed out. It can be seen from the intensity whether the crystal breaks down during the measurement, and from the position of the maximum whether the crystal changes its orientation. After REF (reference reflections) follow the data on these reflections and the criteria that are used to determine whether the orientation should be determined again automatically.

The UB matrix is printed out after SUB (set UB matrix) and the wavelength appears after WAV. We have already discussed the parameters DET and DIS in connection with the automatic orientation determination. The command TLC is followed by the lattice constants, and SDC (start data collection) by the initial values of H, K and L and ΔH. After that comes the output. For each reflection the computer then prints out h, k and l, the azimuth ψ, the number of attenuation filters, the first background, the total intensity and the second background. When reference reflections are measured, the results are printed out on a separate line.

Accurate determination of the lattice constants

The program described earlier for determining the crystal orientation yields lattice constants with an inaccuracy of about 0.2 %. This method cannot give more accurate values because the maximum of the reflection is used instead of the 'centre of gravity'. This causes discrepancies, because the Kα radiation used is not monochromatic, but a doublet of Kα_1 and Kα_2 radiation, which is manifest in a splitting of the reflections at high values of θ.

A special program in which centres of gravity are determined has therefore been designed. This program also makes use of the fact that the rotation about the ω-axis is more accurate than the rotation about the ϕ-axis. The method reduces the inaccuracy in the lattice constants to a value usually less than 0.02 %.

From relation (5), $d^* = 2 \sin \theta$, it follows that d^* can be determined by simply measuring ω, provided $\omega = \theta$, i.e. provided the bisector setting is exact. In

LAT	3	2 0 0	0 1 0	0 0 1			
6	0	0	7·350	8·61	357·93	2580
8	0	0	9·830	8·51	357·92	1561
10	0	0	12·305	8·60	357·91	4149
12	0	0	WK			
14	0	0	WK			
16	0	0	19·890	8·57	357·92	2551
18	0	0	22·505	8·57	357·91	251

D* (2 0 0)= 0·08523   0·023

4	9·841	-0·00011
3	7·375	-0·00024
5	12·329	-0·00010
8	19·948	0·00028
-4	-9·791	-0·00005
-3	-7·325	0·00006
-5	-12·288	0·00024
-8	-19·909	-0·00007

0	2	0	10·670	-78·77	38·48	8604
0	3	0	16·110	-78·76	38·37	3285
0	4	0	21·705	-78·73	38·38	1959

D* (0 1 0)= 0·18466   0·028

4	21·709	-0·00023
3	16·104	0·00015
2	10·663	0·00022
-4	-21·650	0·00014
-3	-16·053	0·00001
-2	-10·604	-0·00030

0	0	4	9·190	7·19	89·01	1644
0	0	5	11·505	7·20	89·01	3216
0	0	6	13·835	7·23	89·01	131
0	0	7	16·180	7·18	89·02	400
0	0	8	18·565	7·19	89·02	1408
0	0	9	20·990	7·19	89·01	1471
0	0	10	23·435	7·17	89·02	1282

D* (0 0 1)= 0·07971   0·005

8	18·597	0·00001
10	23·490	0·00006
9	21·026	-0·00005
5	11·503	-0·00013
-8	-18·589	0·00005
-10	-23·477	-0·00017
-9	-21·014	-0·00003
-5	-11·497	0·00024

Fig. 10. Example of the output in an accurate determination of the lattice constants. The starting command LAT (lattice constants) indicates the number of lines along which the reciprocal lattice constants have to be determined, followed by the first reflection along each of these lines. Each line is explored to find the strong reflections, with which the refining is carried out; WK is printed out for very weak reflections. Then the lattice constant d_1^* (D^* is given in the example) is printed out, followed by the apparent zero point in ω. Next, for a maximum of four strong reflections and their antipodes, the order of the reflection is given, the position of the centre of gravity and the deviation Δ_n with the value calculated from d_1^*.

practice this is never the case. All the reflections along a line through the origin of the reciprocal space, such as the $h00$ reflections, have their bisector setting at the same value of ϕ and χ. This also applies to the reflections with a negative value of h, provided they are measured at negative values of ω and 2θ. This means that a series of reflections of this type can be measured without making any change in ϕ and χ. The inaccuracy then appears as an apparent shift of the zero point of the ω scale, i.e.:

$$d_n* = nd_1* = 2 \sin(\omega_n - \omega_0), \qquad (6)$$

where n stands for h, k or l, in general for the order of the reflection, and where ω_0 represents the zero shift. By determining the centre of gravity ω_n of two or more reflections along this line it is possible to determine d_1* and ω_0 using the method of least squares.

Only the lengths of reciprocal vectors are found in this way. Angles between reciprocal axes are determined from these with the aid of relations such as

$$(d_{110}*)^2 = (d_{100}*)^2 + (d_{010}*)^2 + 2\, d_{100}* d_{010}* \cos \gamma\,*.$$

Here $\gamma*$ is the angle between $a*$ and $b*$. This equation shows a close relationship to the M matrix; the right-hand side is equal to $M_{11} + M_{22} + 2M_{12}$.

The program is started with the command LAT (lattice constants), followed by the number of lines (no more than 9), along which measurements are to be made, and the indices of the first reflection along each line, e.g. 100, 010, 001, 110, etc.; see *fig. 10*. If it is known that there are no $hk0$ reflections with odd h, as in our example, then 200, 210 or 220 can immediately be entered.

An exploration is first carried out for each line. The intensity I for all reflections within the region of θ values given by TRG is determined. The four reflections with the highest value of $I \tan \theta$ are stored for the determination of the centre of gravity. Multiplication by $\tan \theta$ ensures that reflections for small θ are not the only ones taken in acount. During this exploration ϕ and χ are refined. Next, ϕ and χ are left unchanged and the centre of gravity is determined for the four stored reflections, and also for the corresponding reflections with negative n, in so far as $\theta < 50°$ (the lower limit of ω is $-50°$). Finally the best values of d_1* and ω_0 are calculated and printed out, followed by the centre of gravity of each reflection and the deviation Δ_n:

$$\Delta_n = 2 \sin(\omega_n - \omega_0) - nd_1*.$$

From the values of d_1* given in fig. 10 the lattice constants are found in this case by dividing the wavelength $\lambda = 0.7107$ Å by the d_1* values. This gives $a = 16.677$, $b = 3.848$ and $c = 8.916$ Å.

The determination of the centre of gravity ω_n of a reflection profile is based on the defining equation

$$\omega_n = \frac{\int \omega I(\omega)\,d\omega}{\int I(\omega)\,d\omega}.$$

In this equation $I(\omega)$ represents the intensity after correction for the background; the integration must extend over the complete profile. For the computer calculation of the integrals we must decide the range of integration and choose the step size and measuring time per step. The choice of this time depends on the intensity at the maximum. Both accuracy and time taken are of importance in each choice.

We decided on a step size of 0.03°, which is about one-fifth of the width of the reflection profile at half height. The range of integration chosen is the area between the points where the intensity has decreased to 5% of the peak value, and the measuring time was taken as 1000 divided by the intensity of the maximum.

It will simplify the discussion of the errors introduced in this situation if we count ω from the peak ω_m of the profile. In this case

$$\omega_n = \omega_m + \frac{\sum(\omega_i - \omega_m)I_i}{\sum I_i},$$

where I_i is the intensity measured in step i at the angle ω_i. In practice it was found that about 30 steps are needed; the denominator is about 5000 and the numerator is small. The effect of errors due to the counting statistics is greatest when $\omega_i - \omega_m \approx 0.2°$, where $I_i \approx 200$. An error of $\sqrt{200}$ (i.e. about 14) in I_i contributes $0.2 \times 14/5000 \approx 0.0006°$ to ω_n. If some 10 steps give a contribution of this magnitude and the contributions of the other steps are negligible, the contribution of the counting statistics to the standard deviation of ω_n is found to be about 0.002°.

Another source of errors is the assumption that the point where the intensity amounts to 5% of the peak value lies exactly between two measuring points. The last measuring point is taken in full even if it lies close to the 5% limit, in which case it should be halved. The maximum error in ω_n is half the contribution at the last point. If $\omega_i - \omega_m = 0.4°$ and $I_i = 50$, the error is then 0.002°. This error, which may occur at both extremes, is comparable with the error due to the counting statistics.

The error introduced by the uncertainty in the correction for the background is comparable in magnitude, and so also is the maximum error in the ω scale. The absorption by the crystal and the limited accuracy with which the crystal can be aligned also lead to errors of comparable magnitude. We therefore believe that the compromise between speed and accuracy is not far from the optimum.

Summary. The article describes a four-circle single-crystal diffractometer, Philips type PW 1100, which is controlled by a built-in computer with a storage capacity of 8192 words of 16 bits. The lattice constants and orientation of an unknown crystal of arbitrary orientation can be determined in about half an hour. The intensity measurements are also fully automatic. For strong reflections the inaccuracy of the measurement is between 1 and 1.5%; for weak reflections it depends on the counting statistics. The lattice constants can be determined with an accuracy of 0.02%. If necessary, the orientation of the crystal can be redetermined within a few minutes if it alters by more than a specified amount during the intensity measurements. Considerable attention is given to the software, with emphasis on automatic determination of orientation, the automatic measuring process, and the accurate determination of the lattice constants.

'Droplet interferometry' for investigating smooth surfaces

Very small irregularities on smooth surfaces can be observed and investigated with instruments whose operation depends on optical interference [1]. One such instrument is the Linnik interference microscope, in which the state of the surface being investigated is compared with that of a very smooth reference surface that forms part of the instrument. When this reference surface is correctly positioned in relation to the test sur-

magnitude of these irregularities is a measure of the roughness of the surface. A modification of this fringe pattern is obtained if the angle between the two surfaces is such that the field of the microscope is completely filled by a light fringe. This gives a 'contour map' of the surface under investigation. Irregularities show up on the map as closed or nearly closed concentric interference fringes.
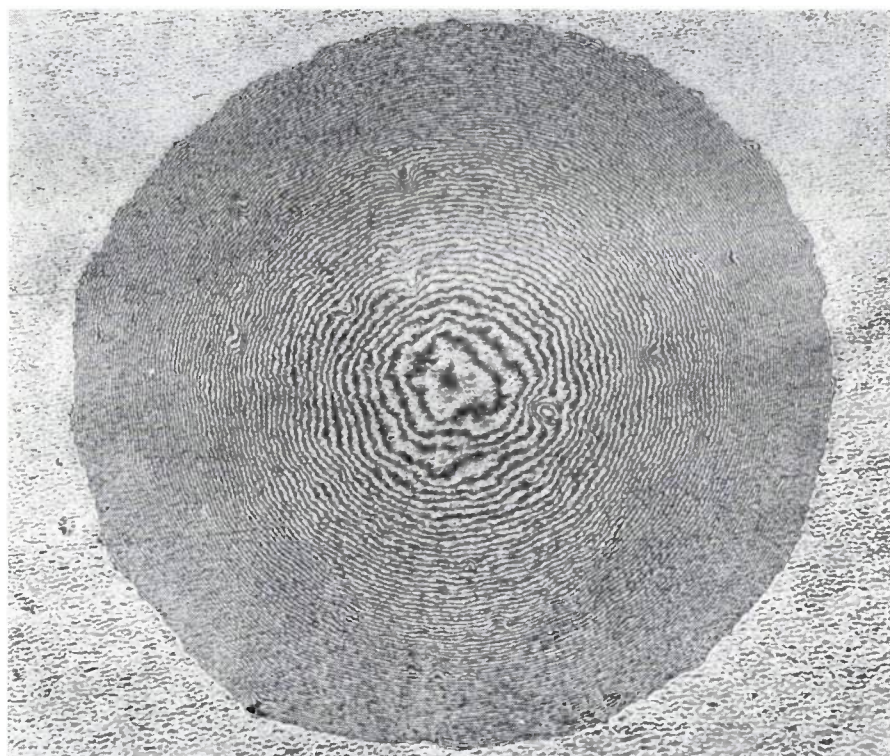
Fig. 1. Interference pattern from a drop of paraffin on a rough surface.

face, reflection of light by the two surfaces gives an interference pattern that can be observed with the microscope. If the test surface is also very smooth the pattern is very regular and consists of straight equidistant light and dark bands ('fringes'). Each successive light fringe corresponds to an increase in optical pathlength of one wavelength, which means that the two surfaces are half a wavelength further apart at each new light fringe. (The unit of measurement is $\lambda/2n$, where λ is the wavelength in vacuum of the light and n the refractive index of the medium between the two surfaces.) If however the surface is not smooth then there are irregularities in the fringe pattern. The

If differences in height of the order of $\frac{1}{2}\lambda$ are to be detected, the reference surface must be extremely smooth and the optical pathlength between the reference and test surfaces should not be too large since this would give less contrast in the fringe pattern [2]. Optical attachments have been developed that enable an ordinary microscope to be used as an interference microscope that satisfies those conditions. The attachment designed by S. Tolansky is an example [3].

[1] S. Tolansky, An introduction to interferometry, 4th edition, Longmans, Green and Co. Ltd., London 1966.

[2] M. Françon, Optical interferometry, Academic Press, New York 1966, pp. 53-56.

[3] S. Tolansky and M. Omar, J. sci. Instr. **30**, 337, 1953.

At Philips Research Laboratories we have recently developed a simple and inexpensive method for measuring irregularities in smooth surfaces that does not require a special attachment. In this method the reference surface is formed by the surface of a drop of a wetting fluid placed on the test surface. The surface of the fluid is by its nature perfect, and in the 'wedge' of liquid at the edge of the drop it is sufficiently close to the test surface for a well defined interference pattern to be produced by the reflection of monochromatic light at the two boundary surfaces (air-liquid and liquid-solid). This is shown in the photographs *figs. 1 to 4*. The spacing of the fringes is determined not only by the location of the liquid surface but also by the refractive index of the liquid. If this is high the spacing of the fringes is smaller for the same angle of contact, giving greater measuring accuracy than with a liquid of low refractive index. Differences in height of less than 0.1 μm can then be measured. The liquid used for the results shown in figs. 2, 3 and 4 was α-bromonaphthalene ($n = 1.658$).

The sense in which the interference fringes are displaced with respect to the edge of the drop indicates whether the irregularity in the surface is a depression or an elevation. The weight of the drop is so small that it does not distort the surface. Since the drop is located *directly* on the surface, the location of the surface with respect to the microscope is not very critical. This means that the method can be used for surfaces that are not completely regular in shape. The area of investigations can be relatively large if the complete edge of the drop is scanned. It is also possible to make this region move slowly across the surface by using a fluid with an appropriate rate of evaporation. An advantage is that the objective can be brought very close to the surface, enabling a high magnification and hence a fine sampling to be used.

<p style="text-align:center">J. P. M. Verbunt</p>

J. P. M. Verbunt is with Philips Research Laboratories, Eindhoven.
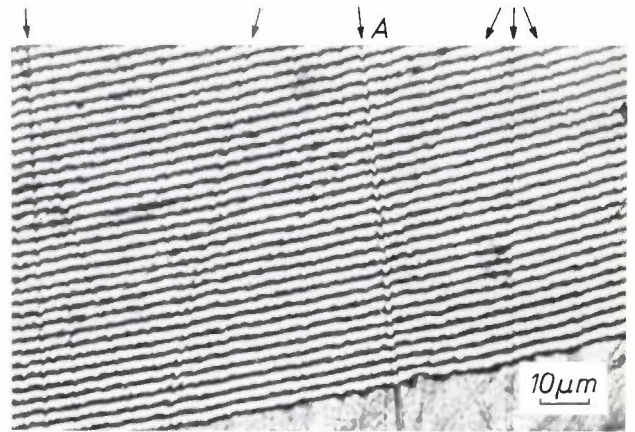
Fig. 2. Quality check by 'droplet interferometry' of a polished ferroxcube surface. The irregularities in the interference pattern are due to polishing scratches (see the arrows). The depth of the scratch at *A* is about 0.1 μm.
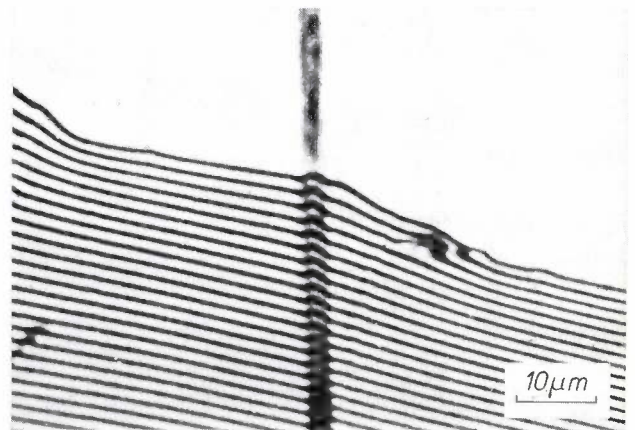
Fig. 3. Interference pattern from the surface of a magnetic head at the gap, where two pieces of ferrite are bonded together with glass. In the polishing process more of the relatively soft material in the gap is removed than of the neighbouring material (about 0.15 μm).
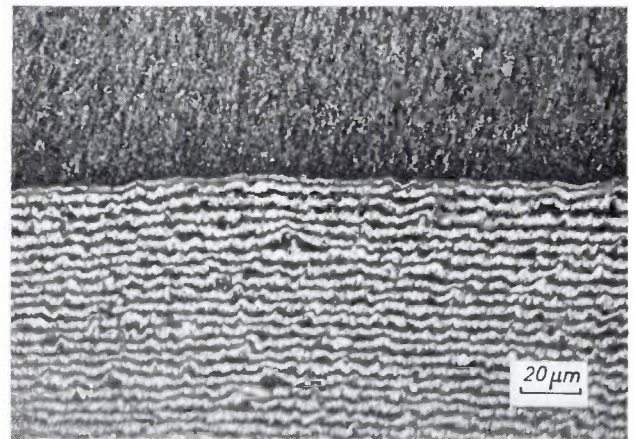
Fig. 4. Interference pattern from the surface of a magnetic tape, used to determine the surface roughness. The distortions of the lines show that the irregularities in the surface are not greater than $\frac{1}{2}\lambda/n$ (about 0.18 μm), and usually much smaller.

Integrated Injection Logic (I²L)

C. M. Hart and A. Slob

In the large-scale integration of logic circuits the basic building blocks must combine great simplicity and compactness with very low dissipation. These requirements are met by the elegant combination of an inversely operated N-P-N transistor with a simple P-N junction for the power supply. In this way 'integrated injection logic' can lead to 'bipolar' integrated circuits whose performance can match that of MOST devices in many respects.

Introduction

The continuous improvement in the manufacture of integrated logic circuits was initially directed at the achievement of shorter delay times. In about 1970, however, the extent to which speeds could be increased reached a practical limit, and the trend is now towards circuits with an increasing number of gates per chip.

For applications where it is desirable to manufacture multi-gate circuits as single units, this large-scale integration (LSI) will constitute an improvement. If LSI is to be a practical proposition, the logic circuits must meet three important requirements. In the first place the basic 'building blocks' must be simple and compact to allow as many of them as possible to be accommodated on a single chip. Secondly, the circuits must be designed in such a way that a reasonable speed does not entail excessive dissipation on the chip. This means that the product τD of the delay time τ and the dissipation per gate D, which is the determining factor here, must be sufficiently low. Thirdly, the fabrication process must be simple and easily controllable and capable of producing a good yield of reliable LSI circuits with as many as a thousand gates. The circuits now in use do not always meet these requirements entirely satisfactorily, and therefore to obtain sufficiently large numbers of gates per chip a new family of logic circuits is necessary.

The circuits that will be described here represent an attractive and novel approach, which meets the requirements [1]. They are in fact multi-collector transistors, i.e. transistors with more than one collector in common base and emitter regions, which are supplied with power by injecting minority carriers into these regions. The injection can take place in two ways: by irradiation with light or by means of a P-N diode.

The first method, which could be of interest in space applications, has the advantage that it requires no external power supply and therefore no supply leads, since small light-actuated current sources are distributed over the whole circuit at the exact places where they are needed. Relatively high levels of illumination are needed, however, if high switching speeds are required.

In the second method the injection is effected by means of a forward-biased P-N diode. This does require an external power supply, but it can be connected into the circuit by a very simple pattern of central power supply rails on the chip. The switching speed can be varied over a wide range by varying the current level. The supply voltage which the circuits require is low (< 1 V). Typically the packing density is 200 gates per mm² and the τD value is 0.7 pJ per gate (the smallest detail in each mask being 7 μm).

The production technique for integrated injection logic circuits ('I²L' circuits) involves only five masking steps. This is rather simpler than the present bipolar production process. Reverting to the normal seven-mask technique, however, it is possible to combine I²L circuits with ordinary integrated circuits (e.g. with analog circuits), thus greatly widening the field of applications.

Injection by light; the photodiode

A P-N junction illuminated with light of a suitable wavelength behaves like a diode with a current source connected to it in parallel (*fig. 1*). Light quanta pene-
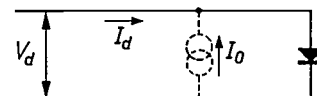
Fig. 1. Illuminated photodiode, symbolically represented by a current source I_0 in parallel with a diode. I_d diode current, V_d diode voltage.

C. M. Hart and A. Slob are with Philips Research Laboratories, Eindhoven.

trating into the diode release electrons and holes, increasing the concentration of both kinds of carrier. The concentration of the minority carriers is increased relatively much more than that of the majority carriers. The extra carriers formed in the neighbourhood of the *P-N* junction are separated by the internal field, giving the effect of a current source [2].

The current I_0 from this source depends on the illumination; its effect on the behaviour of the diode is to displace the normal diode characteristic (curve *1* in *fig. 2*) through a vertical distance I_0 (curve *2*).

In a short-circuited diode ($V_d = 0$) a current $-I_0$ is found in the short-circuit, and in an open-circuited diode ($I_d = 0$) the voltage across the diode is equal to the voltage V_j that is normally required for a current I_0.

The idea is to use such current sources in integrated circuits as distributed supply sources, located at the exact places where they are needed. To make this clear we must first consider the situation found in illuminated transistors.
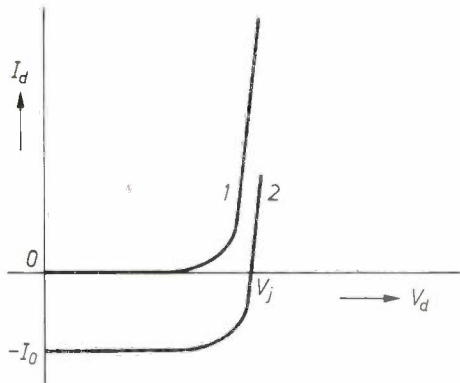
Fig. 2. Characteristics of a non-illuminated diode (curve *1*) and an illuminated one (curve *2*); cf. fig. 1. V_j is the value of V_d at which the current begins to increase steeply.

The illuminated transistor

Fig. 3 illustrates the structure of a phototransistor like those used in our circuits. In principle there are two current sources in this phototransistor: one across the junctions *1-2* and one across the junction *2-3*. In a phototransistor of this type the current source across junction *2-3* gives a much larger current than the one across junctions *1-2*. There are four reasons for this. In the first place, region *3* is much larger than the regions *1*, and consequently the number of carriers generated there is greater. Secondly, the life of the minority carriers before they disappear by recombination is much longer in region *3* than in the regions *1*. Thirdly, the regions *1* are normally completely screened

by the deposited-aluminium connections, which are opaque. Fourthly, owing to the doping gradient, there is usually a 'built-in' electric field in region *2*, causing the minority carriers formed there to drift into region *3*, so that they contribute to the photocurrent of junction *2-3*.
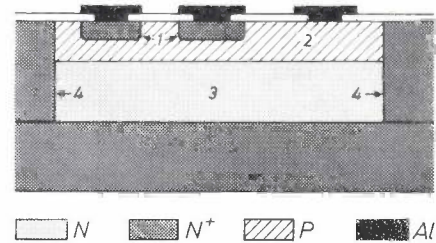
Fig. 3. Schematic cross-section of a phototransistor for integrated injection logic (I²L). On illumination both the junctions *1-2* and the junction *2-3* act as current sources, but the junctions *1-2* only play a subsidiary part since they are largely screened by the aluminium wiring pattern. In the light-activated injection logic described, regions *1* are used as collector and region *3* as emitter. The regions *4* isolate the transistor from its neighbours.

In practice this means that the photocurrent in parallel with the junctions *1-2* can be neglected. To produce transistor action the base-emitter junction must be forward-biased; for the photocurrent to act as the only supply source, the junction *2-3* must be taken as a base-emitter junction, with region *3* as the emitter and the regions *1* as the collector. If we compare this with a conventional planar *N-P-N* transistor, we see that the light-activated transistor is operated 'upside down', for in the conventional version region *1* is the emitter and region *3* the collector. If in these conditions we leave the base unconnected, this region will then become positively charged and a collector, given sufficient voltage, will be capable of drawing a current I_c, where I_c will depend on the surface area of the collector. The base current I_0, as we have seen, depends on the total area of the emitter-base junction. For each collector we can now define a current gain β:

$$\beta = I_c / I_0.$$

In the transistor shown in fig. 3 two collector regions can be seen; more could have been shown since it is a *multi-collector transistor*. The collector regions are of the same size, and since the voltage is the same everywhere in the base region, *each* collector will be capable

[1] See C. M. Hart and A. Slob, IEEE J. SC-7, 346, 1972 (No. 5), and C. M. Hart and A. Slob, 1972 IEEE Int. Solid-State Circuits Conf. Digest tech. Papers, p. 92.
[2] See for example: L. Heyne, Physical principles of photoconductivity, III. Inhomogeneity effects, Philips tech. Rev. **29**, 221-234, 1968.

of drawing a current βI_0. In the application of such an arrangement, the individual collector currents will usually also be approximately equal to I_0. If $\beta > 1$, the collector voltage will then be very low.

This may be explained as follows. Let us consider two identical phototransistors T_1 and T_2; see *fig. 4*. For simplicity, we assume that each of the phototransistors has only one collector and that $I_1 = I_2 = I_0$. The collector of T_1 is connected to the base of T_2, and is used as a switch for short-circuiting or open-circuiting the base of T_2. In such conditions we can construct operating points P and Q in *fig. 5* by intersecting the transistor characteristics with the photodiode characteristic of fig. 2. In this case we have taken the photodiode characteristic upside down, since the positive current direction is reversed. When the base of T_1 is earthed, the current I_0 which the current source supplies to the base of T_1 will leak away through this short-circuit and the base current I_b of T_1 will be zero. Intersection with the photodiode characteristic then gives the operating point Q. When the base of T_1 is open-circuited, however, then I_b is equal to I_0 and the collector of T_1 would then be able to draw a current βI_0. However, this collector current must be supplied by
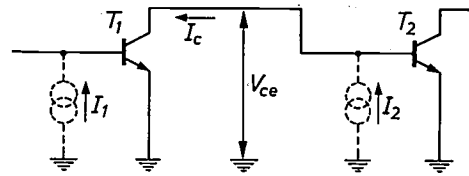
Fig. 4. Standard arrangement for injection logic, in which each collector draws only one base current.
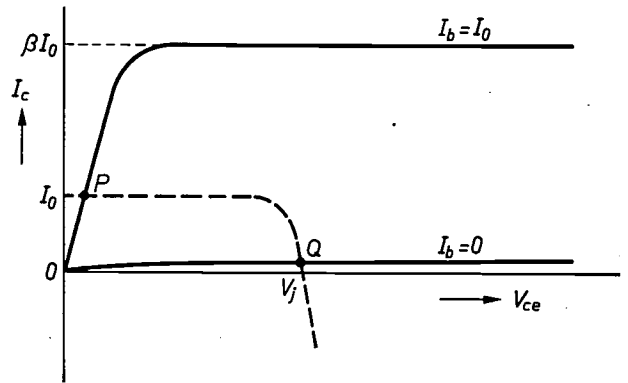
Fig. 5. V_{ce}-I_c diagram for the transistor T_1 from fig. 4 when $I_1 = I_0$. P and Q are possible operating points when the collector is connected to the base of a similar transistor. The point P applies to the case in which the base of T_1 is open-circuited ($I_b = I_0$), and the point Q to the case in which the base is earthed ($I_b = 0$).
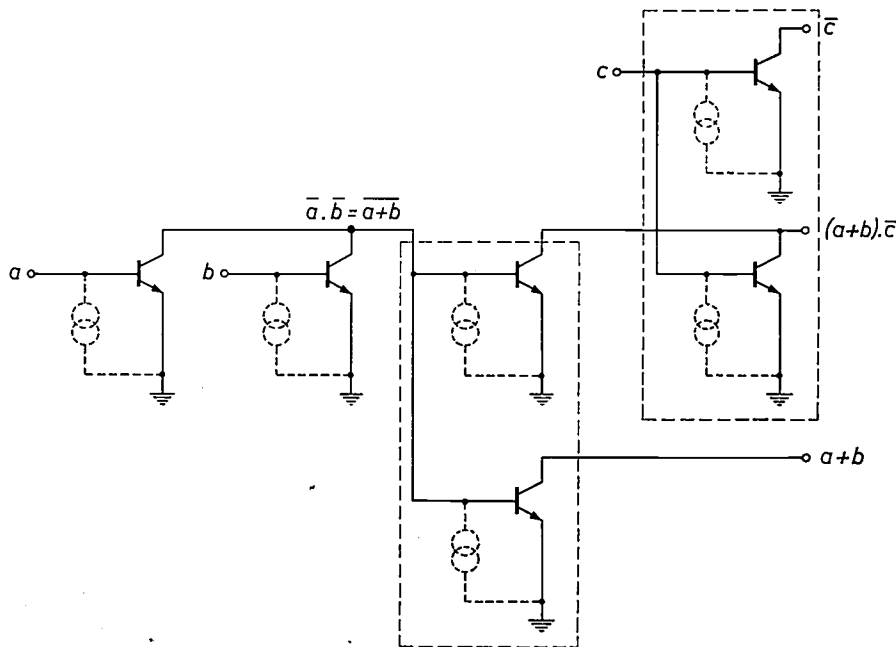
Fig. 6. Example of a logic circuit with transistors (positive logic: the truth of conditions *a*, *b* and *c* corresponds to a high voltage). The dashed lines indicate how the transistors can be combined to form multi-collector transistors.

the current source at the base of T_2, and can therefore be no greater than I_0. This situation then gives the operating point P. In this case V_{ce} of T_1 is very small, which means that the base of T_2 is short-circuited.

Logic circuits with phototransistors

The phototransistors discussed here are very suitable for use in logic circuits. In a circuit like the one of *fig. 6* the transistors produce the logic inversion of the

base signal, while the interconnection of collectors causes the collector signals to perform a 'wired-AND' function. The result is that transistors with interconnected collectors may also be regarded as NOR gates (described in positive logic). As fig. 3 shows, our circuits can easily be given the form of a multi-collector transistor, and mainly because it is desirable to have each collector connected to only one base, the transistors within the dashed lines in fig. 6 can be lumped together to form a circuit as shown in *fig. 7.*
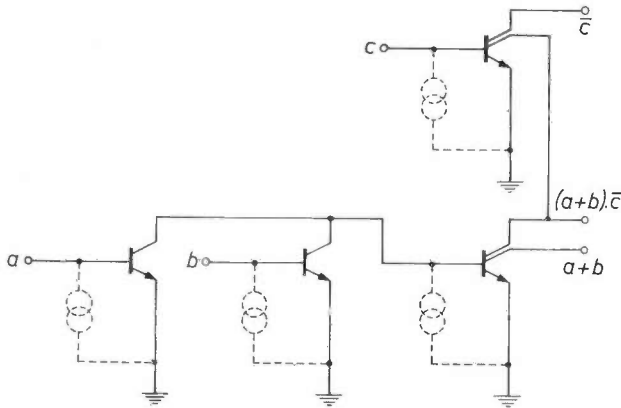
Fig. 7. Injection logic circuit corresponding to the circuit of fig. 6.

Technology

To fabricate a light-injection circuit of the type shown in fig. 3, four masks are required. The process begins with a slice of N^+ material on which a layer of N-type material is grown epitaxially; this corresponds to the layers 2 and 3 in the figure. The first mask is used to make a deep N^+ diffusion 4 in this layer to form the demarcations for the transistors. This diffusion encloses the N-type regions. In the next stage the complete slice is subjected to a P-type diffusion. In this way the bases 2 of the transistors are introduced, and no mask is required because the P concentration is lower than that of the deep N^+ regions. The second mask is now used for introducing the shallow N^+ regions 1, producing the collectors. The third mask is required for generating contact holes in the isolating oxide layer, and the fourth mask produces the final aluminium wiring pattern.

Hole injection by means of a *P-N* junction

In the circuits so far discussed, light was necessary for generating holes in the N-type material of the emitter to produce a current source I_0. This would make the circuits eminently suitable for space applications in which sunlight would be the only source of energy. In the case of illumination by artificial light,

however, the difficulty is that relatively high levels of illumination are required, which are then utilized at relatively low efficiency. This makes it a more economic proposition to inject holes into the emitter region by means of a forward-biased P-N junction. This results in the configuration shown in *fig. 8.* The only difference compared with the circuit in fig. 3 is the extra injector rail 5 now required on the chip. With this power supply system it is easy to select and vary the current level at which the circuits operate.

The fabrication of these circuits requires only one additional masking step, which is the step for generating the separate P-type regions. In some circumstances it is possible to omit the deep N^+ diffusion; *four* masking steps are then sufficient.

Circuit and layout considerations

As can be seen in fig. 8, a multi-collector transistor has the shape of a long rectangle. A suitable and convenient layout for a number of these multi-collector transistors on the chip is to arrange them in linear strips on each side of, and perpendicular to, the injector strip that provides the power supply. This is illustrated in *fig. 9,* which shows two D-type flip-flops and some separate gates. A short injector strip supplies power to three multi-collector transistors and one long rail supplies power to two sets of nine transistors each. Since the relative location of base and collector con-
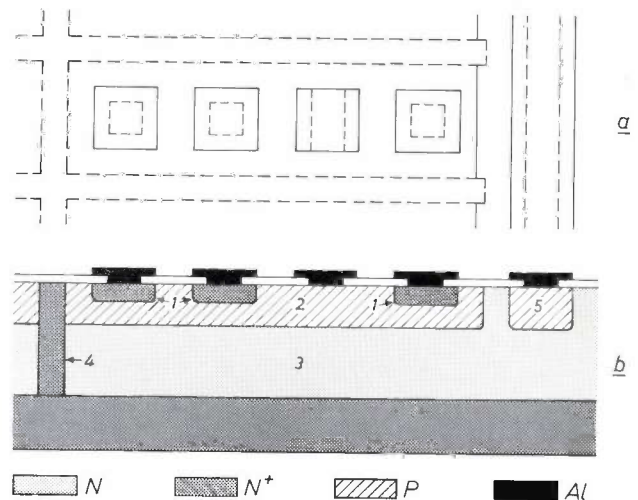
Fig. 8. Schematic cross-section and layout of multi-collector transistors with injection from a forward-biased P-N junction. The figures have the same significance as in fig. 3. Region 5 is an extra P region that with 3 forms the P-N junction for the injection of charge carriers into region 3. The multi-collector transistors are rectangular regions enclosed by the isolating regions 4 (shown dashed in the plan view); they are located in rows perpendicular to the injection rail (region 5), which extends as straight tracks over the whole chip. Compared with the circuit in fig. 3, this configuration requires only one extra masking step for applying the P regions.

tacts is not subject to any fundamental restrictions, they can be laid out in the way most convenient for the wiring pattern required. Space can also be left between collectors where connections can pass a multi-collector transistor without making contact with it, if crossovers are necessary in the wiring pattern. All this is possible provided the multi-collector transistors are not too long.

At high current levels a limit is imposed by the resistance of the base region. This resistance causes a voltage drop across the base, and if this voltage drop is too high the collector currents will not be uniform. Voltage drop across the injector, which would give rise to non-uniform injection, is prevented by applying an aluminium contact rail over the whole length of the injector region. The requirement that the collector
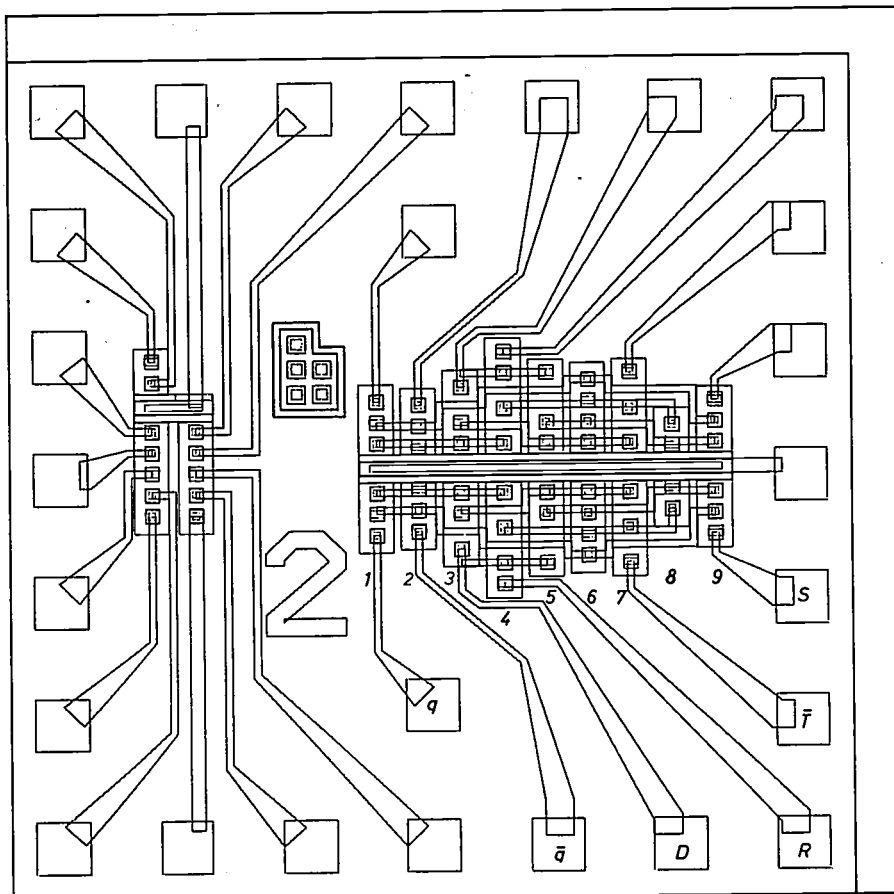
Fig. 9. Layout of a double 'D-type flip-flop' with direct set-and-reset inputs (2×9 transistors), and a few separate gates (3 transistors).

It follows from the geometry that the injected minority carriers are distributed uniformly over each multi-collector transistor. This does not however imply any fixed current level, which the user can in fact select within a wide range.

At low currents a limit is set by the current gain β becoming too small. This situation only arises, however, with currents of the order of nanoamperes. We have already seen that each collector must be capable of drawing one base current. To obtain a suitable operating point (P in fig. 5) this means that the current gain β must be greater than unity. A value of β close to unity is not advisable because of low speed and sensitivity to interference (noise), as we shall see later.

current for the worst-fed transistors should remain greater than the highest values found elsewhere for the base current can be met in a carefully designed circuit at current levels of up to a few tens of milliamperes per chip. This gives the user a freedom of choice of current level over several decades. *Fig. 10* shows a 108-bit shift register consisting of 820 gates, operated at currents ranging from 8 μA to 100 mA. At the low current (about 10 nA per gate) the maximum frequency is 300 Hz, at high currents it is about 800 kHz.

As we saw above, the current level can be selected by the user. This implies that it can be changed during operation. For example, the shift register of fig. 10 has been used in an experimental pocket calculator

that can be given a very low current level during stand-by, i.e. while no operations are in progress and the machine is waiting for a new instruction. This current is automatically switched to a much higher level as soon as the calculator is given an instruction for further operations by pressing one of the buttons. This automatic switch-over makes it possible to keep the average power consumption very low.

A is closed (see fig. 12b). From what has just been said it will then be clear that the base of T_2 is open and T_2 forms a short-circuit for the base of T_3. Transistor T_2 is now 'on' and biased to the operating point P. We now apply at point S a noise voltage V_{S1}, which we represent in fig. 12b by the noise current I_{S1}.

Since T_2 short-circuits the current source I_3, a minimum base current of I_3/β_2 is necessary to maintain this
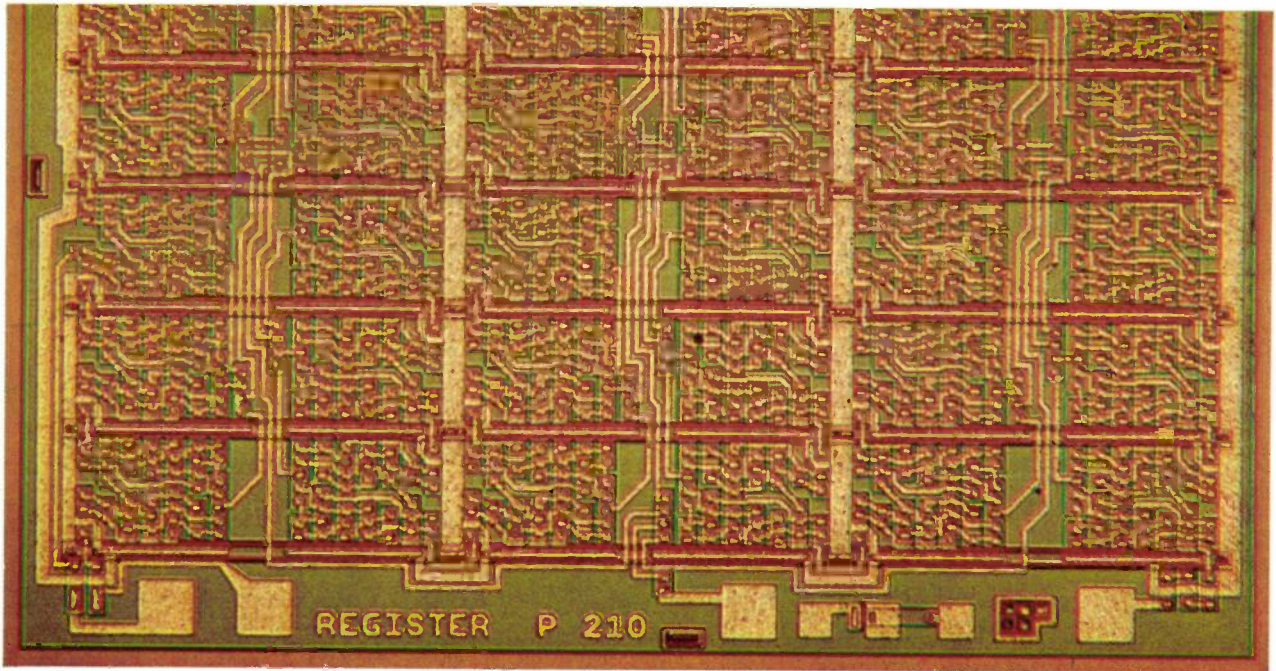
Fig. 10. Part (about half) of a shift register of 108 bits, consisting of 820 gates, used for storing three 9-decimal numbers in an experimental pocket calculator.

Noise sensitivity

The relation between the input voltage V_i and the output voltage V_{ce} of a transistor in the circuit given in fig. 4 can be found with the aid of figs. 2 and 5, or by direct measurement. This relation is shown in *fig. 11.* It can be seen that at the operating point P a relatively small negative noise voltage V_{S1} on the base is sufficient to switch off the collector current, whereas at the operating point Q a relatively large (and positive) noise voltage V_{S0} is permissible before the transistor switches from 'off' to 'on'. The noise margins are perhaps better expressed as current sources.

Let us therefore consider the noise sensitivity at point S in *fig. 12a* for two cases: switch A open and switch A closed. Transistors T_1, T_2 and T_3 do not necessarily have to be identical; there may be differences in geometry, and this is why we have given the transistors different current sources I_1, I_2 and I_3. Furthermore, the current gains β_1, β_2 and β_3 may be different.

We shall first consider the situation in which switch

current in the collector of T_2. Since I_2 is available and T_2 draws no current, the maximum value which the noise source may have is

$$I_{S1} = I_2 - I_3/\beta_2,$$

if the transistor is not to switch from the operating
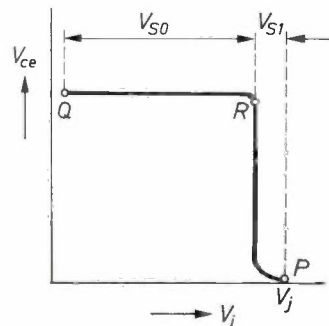
Fig. 11. Curve showing the relation between input voltage V_i and output voltage V_{ce}, e.g. for transistor T_1 in fig. 4. V_{S0} and V_{S1} are the extreme values of noise voltages that are acceptable at the operating points Q and P (see fig. 5) if the logic level is to be correctly reproduced.
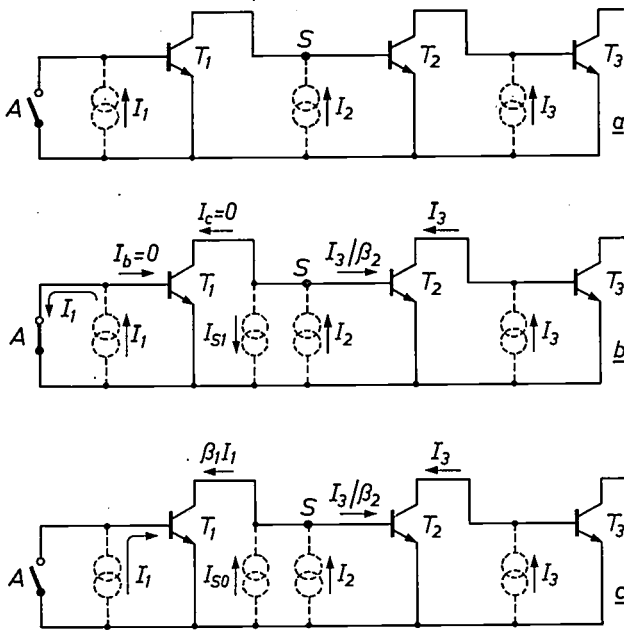
Fig. 12. *a*) Configuration of three transistors $T_{1,2,3}$ with the resultant current sources $I_{1,2,3}$, for calculating the noise sensitivity of an I²L circuit. The noise is assumed to enter at point *S*. *b*) Configuration for calculating the maximum permissible noise I_{S1} for the case where the base voltage of T_1 is zero (switch *A* closed). T_2 is then at the operating point *P* of fig. 5. *c*) The same as (*b*), but for the operating point *Q* (*A* open).

point *P* to point *Q*. The corresponding noise voltage is then given by

$$V_{S1} = \frac{kT}{q} \ln\left(\frac{I_3}{\beta_2 I_2}\right).$$

(*k* is Boltzmann's constant, *T* the absolute temperature and *q* the electronic charge.)

We now consider the situation in which switch *A* is open (fig. 12*c*). Transistor T_1 is now biased to point *P* and short-circuits T_2. Consequently T_2 draws no collector current from T_3, which is 'on'. For a noise source I_{S0} to be able to switch T_2 to 'on', T_2 must first draw the collector current I_3, for which a base current I_3/β_2 is required. In addition, T_1 can draw a current $\beta_1 I_1$ before any substantial increase in its collector voltage appears. It follows that a noise source I_{S0} is acceptable if

$$I_{S0} = \beta_1 I_1 + I_3/\beta_2 - I_2 .$$

This corresponds to a noise voltage

$$V_{S0} \approx V_{\text{j}}.$$

For the situation $I_1 = I_2 = I_3 = I_0$ and $\beta_1 = \beta_2 = \beta_3 = \beta$, we find:

$$I_{S1} = I_0 (1 - 1/\beta); \quad V_{S1} = (- kT/q) \ln \beta$$

and

$$I_{S0} = I_0 (\beta - 1 + 1/\beta); \quad V_{S0} \approx V.$$

In practice there is not likely to be much noise produced on a chip. In our equations, which we obtained by assuming all values of β and all current sources to be equal, a current gain $\beta = 2$ will be quite sufficient. The noise produced in the connections between the chips is expected to be higher, and therefore the first transistor on the chip and the last one on the previous chip are given a larger β and a higher current level. This can be achieved with a suitable geometry.

Switching speeds

When considering the switching speeds of integrated injection logic a distinction must be made between situations of low and high current level. In the first situation only stray capacitances have to be charged and discharged; in the second, charge accumulation in the transistors becomes important, which means that the cut-off frequency is of significance.

At low current levels the propagation delay time per gate (τ) is determined only by junction capacitances and stray capacitances. This delay time will be proportional to the time *t* needed to charge or discharge these capacitances. The time *t* is proportional to $Q/I = CV/I$, and the dissipation *D* is equal to VI, where *V* is the voltage across a forward-biased junction. From this it follows that τD is proportional to CV^2, which is constant.

At high current levels the accumulated charges in the transistors become more important than the charges in the junction capacitance plus the wiring. In this region the transit times are independent of the current level used, since inside the transistor the charges are proportional to I_0/f_t, i.e. proportional to the current divided by the cut-off frequency. The time taken to build up these charges is proportional to the charge divided by the current I_0; the delay time τ is therefore proportional to $1/f_t$ and independent of the current level.

This reasoning explains the measured τD curves. Since the voltage of the injector rail depends only slightly on the current, the dissipation is proportional to I_0. At low currents the product τD is therefore constant, irrespective of I_0. In our circuits it amounts to about 10^{-12} joules, as can be seen from *fig. 13*. At high currents, τ is constant and dependent on the cut-off frequency, which can however, in an inversely operated transistor, as here, be about 20 times lower than with a normally operated one.

Conversion of a logical design into I²L

As we noted earlier, when our circuits are described in positive logic, they represent a negation followed

by a 'wired-AND', which corresponds to a logical NOR.

In general, however, it may be more advantageous to base the design of an I²L circuit not on NORs, but on NANDs. To explain this there are three points that we should consider. First, every collector in our multi-collector transistors represents in effect the negation
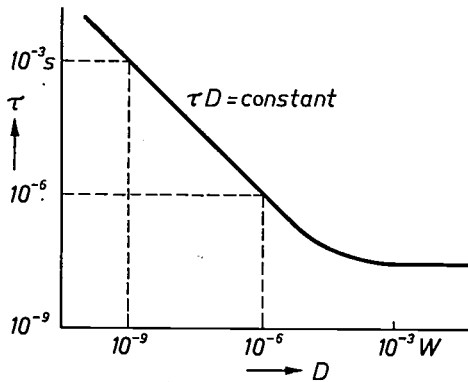
Fig. 13. Relation between delay time τ and dissipation D for a single gate (the τD curve). At low current levels (low dissipation) the product τD is constant; at high currents (high dissipation) τ is independent of the current level.

of the base signal. Secondly, a branching of the signal takes place between base and collectors. Thirdly, each collector or set of interconnected collectors in our circuits can only draw current from one base. It is perhaps more obvious to regard a transistor base, together with the preceding 'wired-AND', as an entity and to think in terms of NANDs when drawing up the design. How this works out in practice will be illustrated by reference to the circuit in *fig. 14*, a D-type flip-flop with direct set and reset inputs.

If for simplicity the symbols of *fig. 15* are introduced for a 'wired-AND' connection and for a multi-collector transistor, the circuit of fig. 14 can be redrawn as shown in *fig. 16*. The six NANDs from fig. 14 directly give the transistors *8, 6, 3, 5, 1* and *2*. The double input $\bar{S}$ required is derived from the additional transistor *9*, fed by S, and the inputs $\bar{R}$ are derived from transistor *4*. The double clock input $\bar{T}$ comes from transistor *7*, which now, however, has to be fed by T. The input D, which goes direct to a 'wired-AND', is regarded as having its associated transistor outside the circuit. The circuit thus requires nine transistors; it is in fact one half of the circuit whose layout is given in fig. 9. The rules for designing a layout of this type will not be dealt with here.

Combination of I²L with other circuits

The circuits described above are made with a five-mask technology starting with an N^+ substrate. This

means that the transistors must always have the emitter connected to the substrate. Transistors used in some other way would also be possible in this technology, but in all cases either the emitter or the collector in N-P-N transistors, or the bases in P-N-P transistor, are connected to the substrate. This design limitation can be removed by starting from a P substrate (see
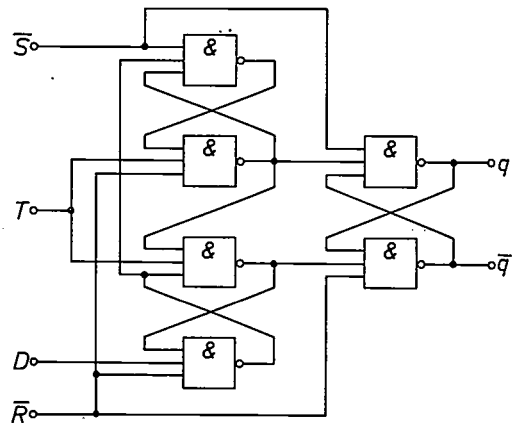
Fig. 14. Logic scheme (in NANDs) of the D-type flip-flop with direct set and reset inputs from fig. 9.
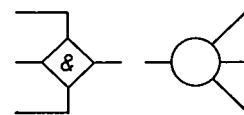
Fig. 15. Symbols for a 'wired-AND' connection and a multi-collector transistor (with three collectors), with which the diagram in fig. 14 can be converted into I²L.
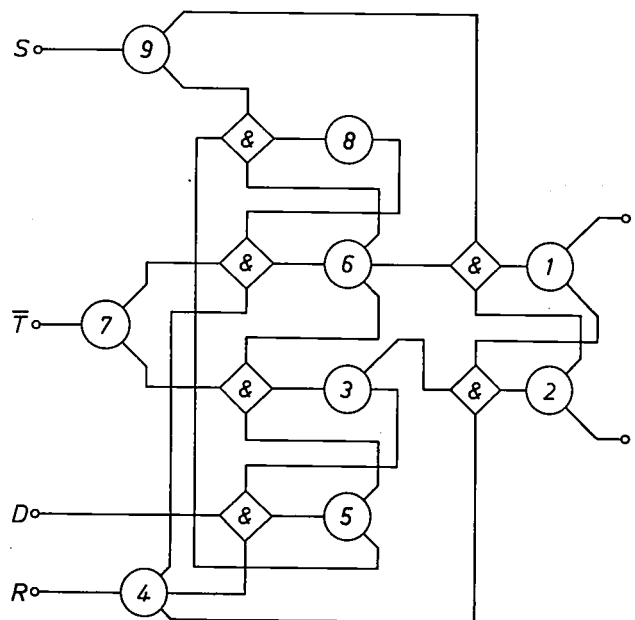
Fig. 16. Conversion of the basic diagram in fig. 14 into a form with multi-collector transistors and 'wired-AND' collector, directly suitable for translation into the layout of fig. 9.

fig. 17), using a buried N^+ layer as auxiliary substrate for the integrated injection logic, and P^+ diffusions for isolating the N-type regions from each other.

In this way, by introducing two additional masking steps (for the buried layer and isolating diffusion) it becomes possible to work with fully isolated transistors, as commonly used in conventional IC technology. Interface circuits of all types can then be produced, and analog circuits can be combined with I²L digital cir-

cuits on the same chip. An example can be seen in *fig. 18*, which shows a 24-bit shift register combined at each of the 24 outputs with a circuit of isolated transistors capable of switching a high a.c. voltage.

Conclusion and results

A convincing proof that I²L is a realistic approach to the LSI problem is our achievement of a packing
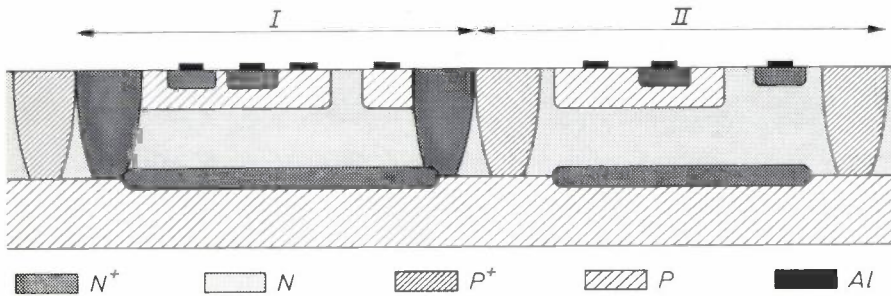
Fig. 17. Combination of injection logic with 'conventional' transistor circuits in a seven-mask technology. The region *I* enclosed by N^+ regions comprises the injection logic. Region *II* shows a 'conventional' transistor. The P^+ regions are isolation diffusions.

N^+ N P^+ P Al
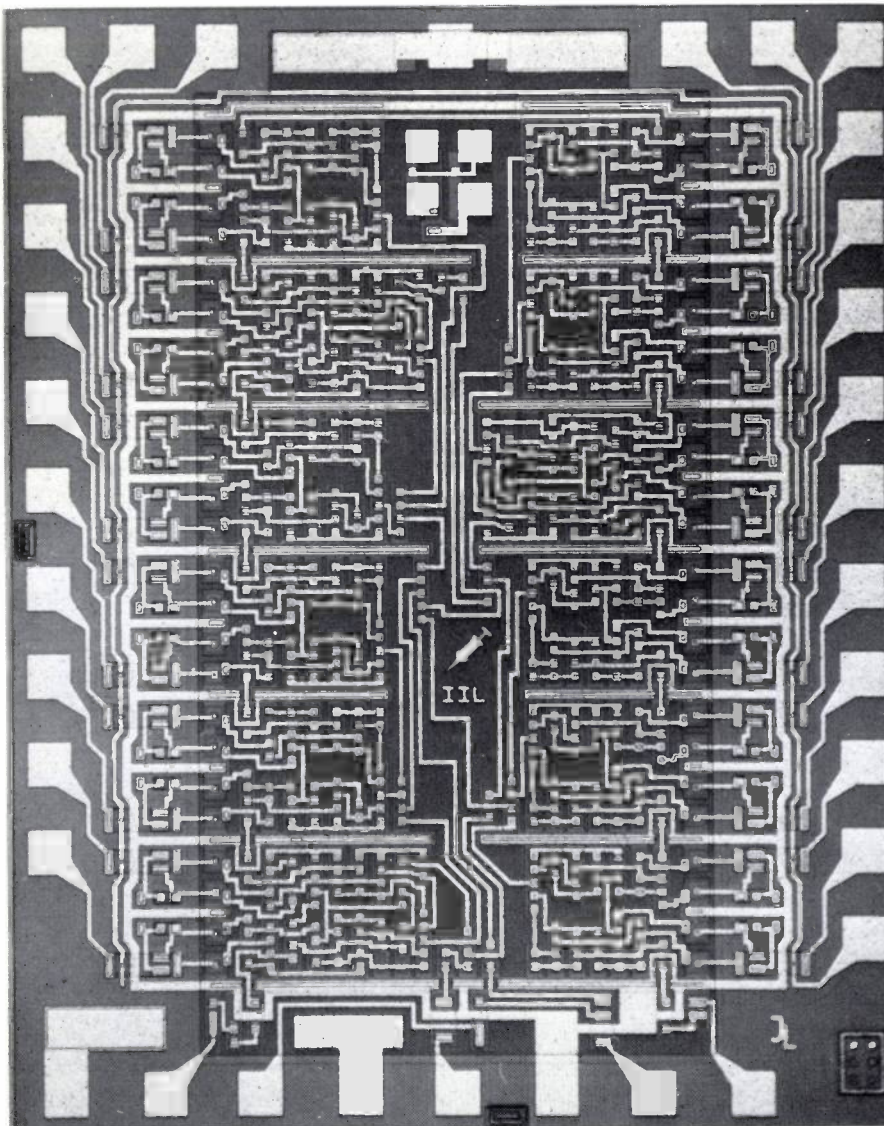
Fig. 18. Example of injection logic combined with analog circuits, showing a 24-bit shift register in which each of the 24 outputs is provided with a circuit of isolated transistors for switching a high a.c. voltage.

Table I. Data for a number of circuits made in I²L. The chip sizes given are overall dimensions, i.e. they include the area taken up by bonding pads and scribing lines. (These were not included in the calculation of the packing density.)

Type of circuit	No. of gates	Chip size (mm)
Tone generator for electronic organ	180	1.2 ×1.5
Liquid-crystal display driver	200	2.2 ×2.4
Counter for digital voltmeter	325	2.2 ×2.4
108-bit shift register	820	2.96×2.85
1536-bit read-only memory	∼ 1000	3 ×4
Control-logic calculator	980	4 ×4

density of 200 gates per mm². The minimum detail was 7 μm, there were five masking steps, and established technology was used throughout. Circuits with up to 1000 gates per chip have been made and put into practical application, e.g. in an experimental pocket calculator.

With a seven-mask technology, normal integrated circuit can be combined with I²L, enabling analog and digital circuits to be produced on the same chip. A wide variety of interface circuits between all kinds of logic can also be achieved. The range of potential applications is therefore very wide. *Table I* gives data for a number of circuits made with the I²L technique (the smallest details were 10 μm). The τD product was between 1 and 2×10^{-12} J.

Summary. Large-scale integration (LSI) requires building blocks that combine great simplicity and compactness with a low dissipation per gate (D), without the speed (delay per gate τ) being too greatly affected. Integrated injection logic (I²L) meets these requirements. I²L circuits consist in practice of multi-collector transistors that are supplied with power by the injection of minority carriers, either by illumination of the base-emitter regions — no external power supply is then required — or from an injector rail that forms a *P-N* junction with the emitter region. The current level of the circuits can be varied in a wide range. I²L circuits with illumination injection can be made with four masks, the others with five. If seven masks are used integrated circuits can be made that consist partly of I²L circuits and partly of isolated conventional transistors. With a minimum detail of 7 μm packing densities of 200 gates/mm² and a τD product of 0.7×10^{-12} J have been achieved. Circuits with 1000 gates per chip have been found to be a practical possibility, e.g. in an experimental pocket calculator.

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, Netherlands   *E*

Mullard Research Laboratories, Redhill (Surrey), England   *M*

Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes (Val-de-Marne), France   *L*

Philips Forschungslaboratorium Aachen GmbH, Weißhausstraße, 51 Aachen, Germany   *A*

Philips Forschungslaboratorium Hamburg GmbH, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany   *H*

MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium.   *B*

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, Netherlands.

R. N. Alcock: Microwave Aircraft Digital Guidance Equipment (MADGE).
SERT J. **6**, 65-68, 1972 (No. 3).   *M*

J. Basterfield, M. J. Josh & M. R. Burgess: Structural features of alloyed contacts to GaAs.
Acta Electronica **15**, 83-87, 1972 (No. 1).   *M*

Chr. Belouet: Kinetics of vapour-phase epitaxial growth of $GaAs_{1-x}P_x$.
J. Crystal Growth **13/14**, 342-345, 1972.   *E*

J. Bloem (Philips Semiconductor Development Laboratory, Nijmegen): Doping in chemically vapour deposited epitaxial silicon.
J. Crystal Growth **13/14**, 302-305, 1972.

P. Blood: The electrical characterization of GaAs using Hall and magnetoresistance measurements.
Acta Electronica **15**, 33-42, 1972 (No. 1).   *M*

R. Boite & J. Neirynck: Analyse des circuits linéaires.
Publ. Gordon & Breach, Paris 1971, xiv + 360 pp.   *B*

J. van den Boomgaard & L. R. Wolff: Growth and properties of the monovariant (Fe,Cr)-$(Cr,Fe)_7C_3$ eutectic system.
J. Crystal Growth **15**, 11-19, 1972 (No. 1).   *E*

G. Bosch: A thermal oscillator using the thermo-electric (Seebeck) effect in silicon.
Solid-State Electronics **15**, 849-852, 1972 (No. 8).   *E*

P. W. J. M. Boumans & G. Brouwer: Studies of photodiodes and phototransistors as detection devices for multichannel emission spectrometry.
Spectrochim. Acta **27B**, 247-255, 1972 (No. 6).   *E*

F. J. A. den Broeder & K. H. J. Buschow: Coercive force and stability of $SmCo_5$ and $GdCo_5$.
J. less-common Met. **29**, 65-71, 1972 (No. 1).   *E*

J. J. van den Broek & W. Staats (Delft University of Technology): The gold-lead system; interstitial dissolution and compound formation.
Scripta metall. **6**, 311-315, 1972 (No. 4).   *E*

H. H. Brongersma & P. M. Mul: Ion scattering: a spectroscopic tool for study of the outermost atomic layer of a solid surface.
Chem. Phys. Letters **14**, 380-384, 1972 (No. 3).   *E*

K. Bulthuis & G. J. Ponsen: On the relaxation of the lower laser level of CO_2.
Chem. Phys. Letters **14**, 613-615, 1972 (No. 5).   *E*

F. M. A. Carpay: Lamellar eutectoid growth.
Acta metall. **20**, 929-934, 1972 (No. 7).   *E*

M. Cathelin: Technologie des diodes Gunn.
Acta Electronica **15**, 69-81, 1972 (No. 1).   *L*

J. Damen, W. Kamminga, J. M. Robertson, C. J. M. Rooymans & J. Schulkes: A furnace for crystal growth at high gas pressures.
J. Crystal Growth **15**, 155-156, 1972 (No. 2).   *E*

M. Davio & J. P. Deschamps: Symmetric discrete functions.
Philips Res. Repts. **27**, 405-445, 1972 (No. 5).   *B*

F. Desvignes & R. Hybois: Thermopiles nouvelles réalisées par des technologies de couches minces.
Techniques Philips 1971, No. 6, 2-20.   *L*

C. Dierieck & Y. Kamp: Polynomial Chebyshev approximation of a complex transfer function.
Computing **8**, 335-342, 1971 (No. 3/4).   *B*

J. A. W. van der Does de Bye & A. T. Vink: Minority carrier lifetime in p-type gallium phosphide.
J. Luminescence **5**, 108-116, 1972 (No. 2).   *E*

W. G. Essers: Metal transfer in MIG welding.
Conf. Int. Inst. of Welding, Toronto 1972, Doc. Comm.
XIV, pp. 1-23.   *E*

E. Fabre & C. Schiller: Caractérisation optique et cristallographique de l'arséniure de gallium.
Acta Electronica 15, 55-67, 1972 (No. 1).   *L*

P. Guétin & G. Schréder: Quantitative aspects of the tunneling resistance in n-GaAs Schottky barriers.
J. appl. Phys. 42, 5689-5698, 1971 (No. 13).   *L*

H. Haug & K. Weiss: On the hydrodynamics of condensed Bose-systems.
Physics Letters 39A, 255-256, 1972 (No. 4).   *E*

H. Haug & K. Weiss: A modified theory of the Kapitza resistance.
Physics Letters 40A, 19-21, 1972 (No. 1).   *E*

L. Hollan, B. C. Easton & J.-M. Le Duc (RTC La Radiotechnique-Compelec, Caen): Elaboration de couches épitaxiales de GaAs pour dispositifs Gunn.
Acta Electronica 15, 11-31, 1972 (No. 1).   *L, M*

F. N. Hooge: Discussion of recent experiments on $1/f$ noise.
Physica 60, 130-144, 1972 (No. 1).   *E*

F. D. Hughes: The characterization of epitaxial GaAs using Schottky barriers.
Acta Electronica 15, 43-53, 1972 (No. 1).   *M*

C. G. J. Jansen & A. J. Verdonschot: The contribution of X-ray photoelectrons to the retarding-field characteristic of a thermionic emitter.
Philips Res. Repts. 27, 454-485, 1972 (No. 5).   *E*

W. H. de Jeu & C. J. Gerritsma: Electrohydrodynamic instabilities in some nematic azoxy compounds with dielectric anisotropies of different sign.
J. chem. Phys. 56, 4752-4755, 1972 (No. 10).   *E*

W. H. de Jeu, C. J. Gerritsma & Th. W. Lathouwers: Instabilities in electric fields of nematic liquid crystals with positive dielectric anisotropy: domains, loop domains and reorientation.
Chem. Phys. Letters 14, 503-506, 1972 (No. 4).   *E*

W. H. de Jeu, C. J. Gerritsma, P. van Zanten & W. J. A. Goossens: Relaxation of the dielectric constant and electrohydrodynamic instabilities in a liquid crystal.
Physics Letters 39A, 355-356, 1972 (No. 5).   *E*

H. Kalis & J. Lemmrich: Frequenzanaloge Prozeßgrößendarstellung, III. Frequenzanaloge Drehzahlregelungen.
Elektrotechn. Z. B 24, 325-328, 1972 (No. 13).   *H*

R. J. Klein Wassink: Continuous hot pressing of reactive materials.
High Temp. - High Pressures 3, 411-418, 1971 (No. 4).   *E*

M. Klinck: Ein digitaler Einzelregler für frequenzanaloge Meßwerte. (Digitale Prozeßgrößendarstellung, IV.)
Elektrotechn. Z. B 24, 328-330, 1972 (No. 13).   *H*

P. Kramer & L. J. van Ruyven (Philips Semiconductor Development Laboratory, Nijmegen): Position of the band edges of silicon under uniaxial stress.
Appl. Phys. Letters 20, 420-422, 1972 (No. 11).

P. Kramer & L. J. van Ruyven (Philips Semiconductor Development Laboratory, Nijmegen): The influence of temperature on spreading resistance measurement.
Solid-State Electronics 15, 757-766, 1972 (No. 7).

J.-P. Krumme, J. Verweel, J. Haberkamp, W. Tolksdorf, G. Bartels & G. P. Espinosa: Thermomagnetic recording in thin garnet layers.
Appl. Phys. Letters 20, 451-453, 1972 (No. 11).   *H*

F. A. Kuijpers: Influence of a magnetic field on the hydrogen equilibrium pressure of RCo_5H_x.
Physics Letters 39A, 240-242, 1972 (No. 3).   *E*

G. Landvogt: Frequenzanaloge Prozeßgrößendarstellung, II. Schwingende Blattfeder als frequenzanaloger Meßgrößenaufnehmer.
Elektrotechn. Z. B 24, 259-261, 1972 (No. 11).   *H*

J. B. van Laren: Het elektronendichtheidsprofiel in een lagedrukmetaaldamp-edelgas-ontlading.
Polytechn. T. Elektr. 27, 233-238, 1972 (No. 7).   *E*

R. Memming & F. Möllers: Studies of the rearrangement-energy in electrode kinetics at SnO_2-electrodes.
Ber. Bunsen-Ges. phys. Chemie 76, 475-481, 1972 (No. 6).   *H*

D. Meyer-Ebrecht: Frequenzanaloge Prozeßgrößendarstellung, I. Meßumformer für frequenzanaloge Instrumentierungssysteme.
Elektrotechn. Z. B 24, 243-246, 1972 (No. 10).   *H*

D. Meyer-Ebrecht: Fast amplitude control of a harmonic oscillator.
Proc. IEEE 60, 736, 1972 (No. 6).   *H*

A. C. Miller, M. R. Hoare, J. D. Laird, A. C. Arntzenius, P. G. Hugenholtz (all with Rotterdam Medical Faculty) **& W. Rey:** Two approaches to arrhythmia detection by digital computer in the coronary care unit.
Folia med. neerl. 14, 209-217, 1971 (No. 5/6).   *B*

R. F. Mitchell & N. H. C. Reilly (Queen Mary College, London): Equivalence of δ-function and equivalent-circuit models for interdigital acoustic-surface-wave transducers.
Electronics Letters 8, 329-331, 1972 (No. 13).   *M*

A. G. van Nie: Oscillator circuit for a vibrating capacitor driven by an RF electric field.
IEEE Trans. IM-21, 153-161, 1972 (No. 2).   *E*

M. Noé: The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics.
Ann. math. Stat. 43, 58-64, 1972 (No. 1).   *B*

S. G. Nooteboom (Institute for Perception Research (IPO), Eindhoven): Production and perception of vowel duration. A study of durational properties of vowels in Dutch.
Thesis, Utrecht 1972.

M. Otala: Circuit design modifications for minimizing transient intermodulation distortion in audio amplifiers.
J. Audio Engng. Soc. 20, 396-399, 1972 (No. 5). *E*

J. A. Pals: A general solution of the quantization in a semiconductor surface inversion layer in the electric quantum limit.
Physics Letters 39A, 101-102, 1972 (No. 2). *E*

J. A. Pals: Experimental verification of the surface quantization of an *n*-type inversion layer of silicon at 300 and 77 °K.
Phys. Rev. B 5, 4208-4210, 1972 (No. 10). *E*

J. A. Pals: Quantization effects in semiconductor inversion and accumulation layers.
Thesis, Eindhoven 1972. *E*

R. F. Pearson: Ferroelectrics and their applications.
Physics Bull. 23, 360, 1972 (June). *M*

Ph. Piret: High-rate binomial convolutional codes.
Philips Res. Repts. 27, 446-453, 1972 (No. 5). *B*

J. Polman & P. C. Drop: Electron temperature and Hg(6³P) atom density decay in a mercury — rare-gas afterglow.
J. appl. Phys. 43, 1577-1580, 1972 (No. 4). *E*

W. Radziwill: Theorie des Betriebsverhaltens kleiner elektronisch kommutierter Gleichstrommaschinen mit Drehzahlregelung durch Variation der Pulsbreite und Pulsphase der Ankerspannung.
Ingenieur 84, E 41-49, 1972 (No. 21). *A*

H. Rau & T. R. N. Kutty: *PVT*-Messungen an konzentrierter Salzsäure bis 500 °C und 1500 bar.
Ber. Bunsen-Ges. phys. Chemie 76, 645-651, 1972 (No. 7). *A*

J. Revuz, J.-P. Michel & J. Michel: Senseur d'horizon à bolomètre et modulateurs électro-optiques.
Acta Electronica 15, 101-130, 1972 (No. 2). *L*

T. E. Rozzi: Equivalent network for interacting thick inductive irises.
IEEE Trans. MTT-20, 323-330, 1972 (No. 5). *E*

B. Schneider: Effect of crystalline anisotropy on the magnetostatic spin wave modes in ferromagnetic plates, I. Theoretical discussion for infinite plates.
Phys. Stat. sol. (b) 51, 325-338, 1972 (No. 1). *H*

G. Schréder: Contribution à l'étude de l'effet tunnel dans les semiconducteurs.
Thesis, Paris 1972. *L*

K. Schünemann & B. Schiek: Noise in cavity-stabilized microwave oscillators.
Philips Res. Repts. 27, 486-507, 1972 (No. 5). *H*

E. Schwartz (Technische Universität, Braunschweig) **& H. Bex:** Grenzen für die Beträge der Streumatrixelemente von rotationssymmetrischen passiven Dreitoren, insbesondere Zirkulatoren.
Arch. Elektronik & Übertr.technik (AEÜ) 26, 336-342, 1972 (No. 7/8). *A*

M. Sintzoff: Calculating properties of programs by valuations on specific models.
Proc. ACM Conf. on proving assertions about programs, Las Cruces, N.M., USA, 1972, pp. 203-207. *B*

C. G. Sluijter: Televisie in het natuurwetenschappelijk onderzoek.
Onderwijsinformatie No. 35, 20-25, 1972. *E*

T. L. Tansley: Temperature dependence of hole diffusion length from spectral response of Au-n⁺ GaAs photodiodes.
J. Physics D 5, 1146-1152, 1972 (No. 6). *M*

R. Veilex: Introduction *(to Gunn-effect issue of Acta Electronica)*.
Acta Electronica 15, 7-8 (in French), 9-10 (in English), 1972 (No. 1). *L*

P. H. J. Th. Ververgaert, P. F. Elbers (both with Centrum voor Submicroscopisch Onderzoek, Utrecht), **A. J. Luitingh & H. J. van den Berg:** Surface patterns in freeze-fractured liposomes.
Cytobiologie 6, 86-96, 1972 (No. 1). *E*

J. D. Wasscher & J. Dieleman: Anisotropy of the optical constants of GaSe near the band edge.
Physics Letters 39A, 279-280, 1972 (No. 4). *E*

H. W. Werner & J. v. d. Berg: Determination of the ion-induced electron yield from a statistical treatment of pulse-height distribution curves measured with an ion-electron-scintillation detector.
J. appl. Phys. 43, 1524-1529, 1972 (No. 4). *E*

H. W. Werner, A. Venema & A. J. Linssen: Design of a small magnetic-sector mass spectrometer with reduced mass discrimination.
J. Vacuum Sci. Technol. 9, 216-219, 1972 (No. 1). *E*

J. S. C. Wessels & M. J. F. van Leeuwen: Photosystem-II activity in subchloroplast fragments prepared by the action of digitonin.
Energy transduction in respiration and photosynthesis, Proc. Conf. Bari 1971, pp. 537-550; 1972. *E*

F. F. Westendorp: Domains in SmCo₅ at low temperatures.
Appl. Phys. Letters 20, 441-443, 1972 (No. 11). *E*

H. J. de Wit & C. Crevecoeur: The electrical conduction of As₂Se₃ glass at high fields.
Solid-State Electronics 15, 729-738, 1972 (No. 7). *E*

H. Zijlstra: Magnets that attract hydrogen.
Chem. Technol. 2, 280-284, 1972 (May). *E*

A flexible method for automatic reading of handwritten numerals

M. Beun

I. General description of the recognition method

II. Thinning procedure and determination of the special points

For the last twenty years or so, efforts have been made in various parts of the world to devise automatic methods of reading handwriting, so that large volumes of handwritten data for clerical processing can be fed direct to a computer. The present article describes an automatic method of reading numerals. In an experimental equipment using this method, which is based on a few simple principles, more than 91% of a collection of 10 000 numerals written by hand without any restriction were read correctly, some 3% were read wrongly and 6% were rejected as unrecognizable. Since the numerals were not written particularly well (they were obtained in a door-to-door survey) this result compares very well indeed with the results of other methods, and in fact the flexibility of the method allows further considerable improvement. The author comes to the conclusion, however, that the reliability in the automatic reading of forms with many numerals written on them (such as giro cards) will never be completely adequate unless the people who have to fill in the forms are prepared to give a little cooperation. In part I of the article the recognition system is described in general terms; in part II [*] the author deals at greater length with some important details of the procedure.

I. General description of the recognition method

Introduction

Information to be fed into a computer is frequently presented in the form of handwriting. A familiar example in some countries is the giro card, which is filled in by the account holder, usually by hand. The problem is by no means limited, however, to the giro service or to the cheque and payment-transfer departments of banks. There are very many other applications of the computer, both administrative and scientific, in which the input consists of large quantities of handwritten data. At the present time the data are usually processed by hand to bring them into a form in which the computer can understand them. In the giro service, for

example, many hundreds of punch operators are occupied in this kind of work. Mark sensing is another system that has been used, in which marks made in predetermined positions on documents can be directly read by the computer, either optically or magnetically. Filling in these marks presents too many problems, however, and they are also difficult for people to read back. There is clearly a need for a machine that can put data into a computer by reading handwriting directly.

For most practical purposes numerals are adequate. Giro services, for example, can manage with the

---

Ir M. Beun is with Philips Research Laboratories, Eindhoven.

amount and two account numbers. For many other applications the information can be presented in a numerical code. In some countries the post office now asks people to add a numerical 'postcode' for the town to the address on the envelope. The considerable research that has already been done on optical character recognition has therefore usually been limited to

this image is electronically transformed into a matrix of occupied and unoccupied positions (*fig. 2*). The occupied positions (black in fig. 2) reproduce the numeral, the unoccupied positions (grey) reproduce the surroundings. This matrix can easily be stored in the memory of a computer and is therefore very suitable for electronic processing.
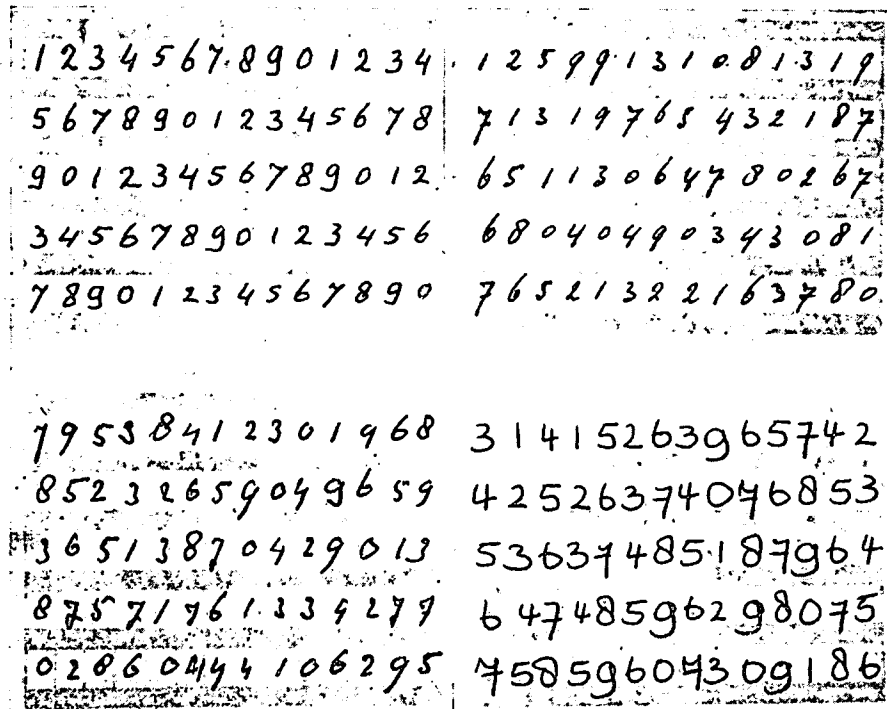
Fig. 1. Example of the test material used in our experiments. Test subjects were asked to write any 70 numbers on a sheet of paper with printed lines enclosing spaces measuring 5 by 7 mm. Four of these sheets are shown here. The first was filled in by the author himself with carefully written numerals. These were read with no errors using the method described here and a very simple recognition procedure.

numerals, and even in many cases to stylized numerals. In our research we have concentrated on the automatic recognition of non-stylized numerals, our aim being to impose no restrictions whatever on the handwriting. The principle of our system will now be described, and it will then be compared with the considerable work by others in this field.

Principle of the recognition method

Fig. 1 gives an idea of the types of numerals used in our experiments. The only rule laid down is that the numerals must be written on a preprinted form, with no more than one numeral per space. Before applying the recognition procedure the numerals are put into a form in which they can readily be processed electronically. This is done by means of a television camera. The camera produces an image of the numeral, and

No knowledge of computers is required for the understanding of this article. The reader need only think of a drawing of a matrix in which occupied and unoccupied positions are distinguished from each other in some way; for example by making the occupied positions black and the others white.

Very briefly, the recognition process takes place in the following phases. First the numeral is 'thinned' to a skeleton (skeletonized), i.e. reduced to a pattern that is nowhere thicker than one square. Next the end points and forks or junctions in this pattern are found, and then the numeral which the camera has read is determined from the shape of the skeleton, and in particular from the relative positions of these special points. In this last phase, the actual recognition process, a procedure is used that is drawn up after studying the characteristics of a large number of numerals written by test subjects. By continuously

checking this recognition procedure against numerals written by other people, and if necessary introducing new criteria, the accuracy can be steadily improved. This is what gives our method its flexibility [1].

The operations described can all be performed by a computer. Working with a computer has great advantages in the development of the recognition method, since it enables a quick test to be made of the effect of changes in the thinning procedure or in the recognition scheme. With a final version of the method and development of a machine for practical use, it could well be advantageous to design a special electronic system capable of performing the necessary operations faster than a computer. Since we are for the present only concerned with the development of the method, these questions fall outside the scope of this article.

Among the recognition methods that others have developed [2] there are some that work as follows. First, the numeral to be read is enclosed in the smallest possible rectangle, which is then divided up into a grid of smaller rectangles, and some of the character-

line sections of a number of reference numerals. Next, by applying certain criteria, the figure is determined from the correlations thus found. Some of these methods are not very flexible and they are therefore generally used for stylized, or at least 'well written' numerals. The basis is usually the 'thick' numeral [3], but various methods of skeletonization have been reported, not only of handwritten numerals [4] but also of chromosomes [5]. In two cases an attempt has been made, in much the same way as in our method, to use the special points of the skeleton for the recognition of handwritten characters [6]; as far as we know, however, these investigations have not led to practical results. A machine for reading non-stylized numerals [7] has been developed in Japan; this is used for reading postcodes on letters. These postcodes consist of *three* numeral digits, and the machine reads 70% correctly and 2% wrongly, while 28% are rejected as unrecognizable.

The good results that we have obtained with our method are in no small part due first to the applications of a special thinning procedure that makes use of
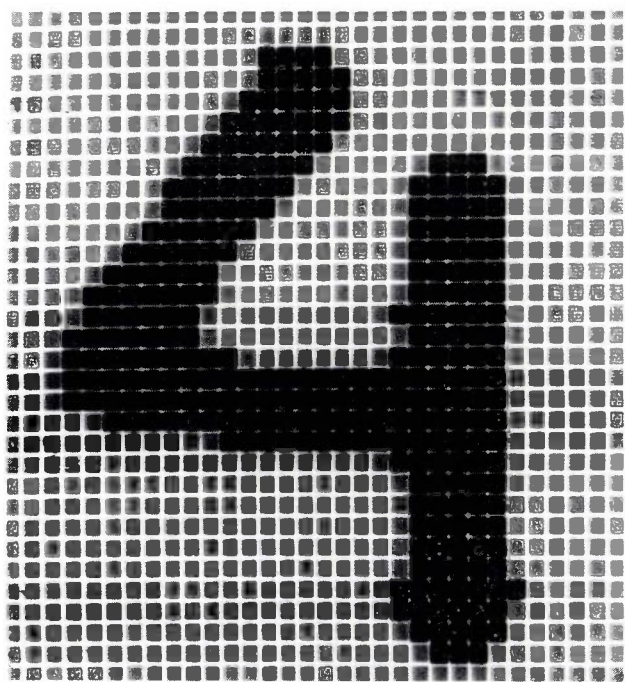
Fig. 2. Matrix of occupied (black) and unoccupied (grey) positions, the occupied positions forming the numeral to be recognized. A matrix of this type is very suitable for storage in a computer memory and further electronic processing.

istics of the line sections of the numeral in the various spaces of the rectangular grid are determined. These characteristics (e.g. the direction of the line section, or the curvature, or the presence of a junction) are compared with the characteristics of the corresponding

[1] In this article it will be seen that in addition to the skeleton we sometimes had to use the unthinned numeral; most of the criteria that were then applied were derived from the work of E. de Boer and F. L. A. M. Thissen of this laboratory.

[2] P. Saraga, J. A. Weaver and D. J. Woollons, Optical character recognition, Philips tech. Rev. **28**, 197-202, 1967.
G. Nagy, State of the art in pattern recognition, Proc. IEEE **56**, 836-862, 1968.

[3] E. C. Greanias, P. F. Meagher, R. J. Norman and P. Essinger, The recognition of handwritten numerals by contour analysis, IBM J. Res. Devel. **7**, 14-21, 1963.
R. Bakis, N. M. Herbst and G. Nagy, An experimental study of machine recognition of hand-printed numerals, IEEE Trans. **SSC-4**, 119-132, 1968.
I. Sheinberg, The INPUT 2 Document Reader (A new optical character recognition system), Pattern Recog. **2**, 167-173, 1970.
D. J. Quarmby and J. Rastall, Experiments on handwritten numeral classification, IEEE Trans. **SMC-1**, 331-338, 1971.
J. T. Tou and R. C. Gonzalez, A new approach to automatic recognition of handwritten characters, Proc. Two-dimensional Digital Signal Processing Conf., Columbia, Missouri, 1971, pp. 3-2-1 to 3-2-10.

[4] P. Saraga and D. J. Woollons, The design of operators for pattern processing, I.E.E. N.P.L. Conf. on Pattern Recognition, Teddington 1968, pp. 106-116.
E. E. Triendl, Skeletonization of noisy handdrawn symbols using parallel operations, Pattern Recog. **2**, 215-226, 1970.
E. S. Deutsch, Thinning algorithms on rectangular, hexagonal, and triangular arrays, Comm. ACM **15**, 827-837, 1972 (No. 9).

[5] C. J. Hilditch, Linear skeletons from square cupboards, Machine Intelligence 4, ed. B. Meltzer and D. Michie, Edinburgh Univ. Press 1969, pp. 403-420.

[6] H. Sherman, A quasi-topological method for the recognition of line patterns, Information Processing, Proc. int. Conf. Paris 1959, pp. 232-238.
S. Tomita, S. Noguchi, J. Oizumi and K. Iwamoto, Recognition of handwritten Katakana characters, Electronics & Comm. in Japan **50**, No. 4, 174-183, 1967.

[7] H. Genchi, K. Mori, S. Watanabe and S. Katsuragi, Recognition of handwritten numerical characters for automatic letter sorting, Proc. IEEE **56**, 1292-1301, 1968.
K. Mori, H. Genchi, S. Watanabe and S. Katsuragi, Microprogram controlled pattern processing in a handwritten mail reader-sorter, Pattern Recog. **2**, 175-185, 1970.

the work of P. Saraga and D. J. Woollons [8] and secondly to the flexibility built into it by allowing new criteria to be added continuously to the recognition procedure.

The description of our method will now be continued, followed by an examination of the provisional results obtained in an experimental arrangement used for the recognition of several thousand numerals. Part I of the article includes an evaluation of these results, plus the findings of a small experiment designed to find out how accurately *people* could read the numerals.

1) break points, i.e. points that leave a gap in the line when they are removed;
2) end points, i.e. points that have only one neighbour;
3) loop points, i.e. points that give a loop if they are removed (the loop is formed by the four neighbouring points).

A skeleton therefore consists exclusively of break points, end points and loop points. The loop points are fairly rare; an example is the crossing point in fig. 5b.

Making a good skeleton is not so easy as it may seem, since the shape of the skeleton depends to a great extent on the thinning procedure adopted. We need a
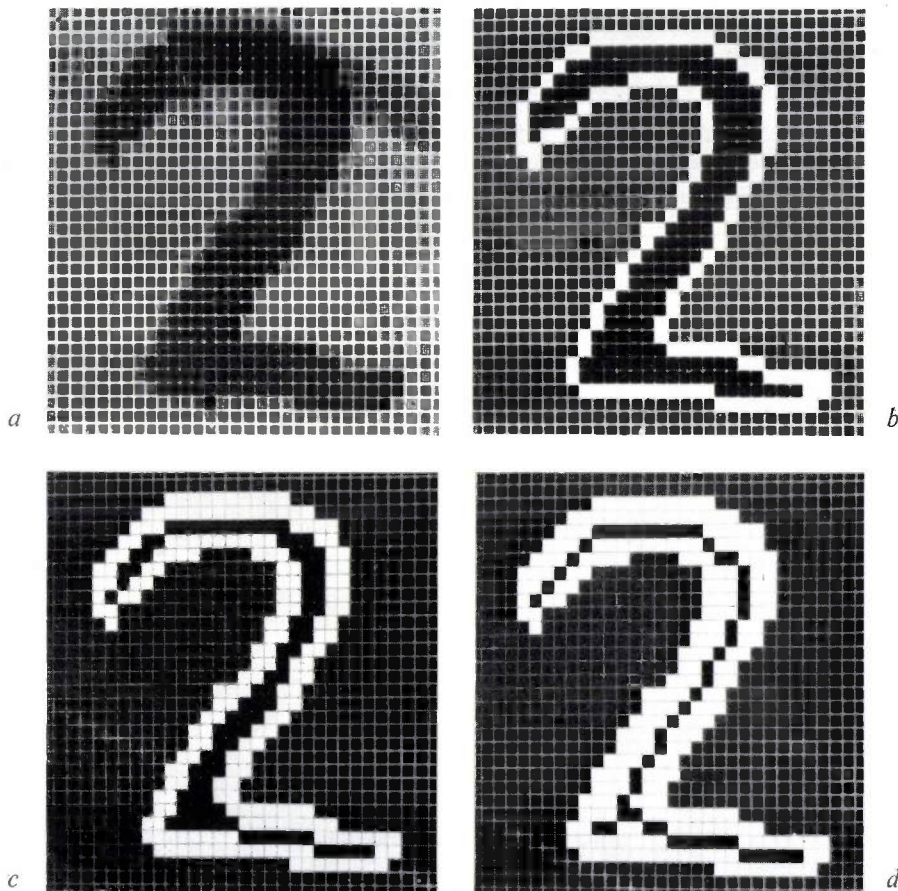
Fig. 3. The skeletonization process. The pattern of positions forming the figure (a) is thinned in a number of steps (b and c) by removing points around the edge of the numeral. Points that must not be removed are end points and points whose removal would cause a break in the pattern. The thinning process is continued until there are no more points that can be removed (d). The figure then remaining is the skeleton.

Skeletonization

We start from the situation where the numeral is present in the form of a matrix with occupied and unoccupied positions. *Fig. 3* shows how a skeleton is made by successively removing points on the edge of the original shape. This thinning process is continued until the skeleton only contains points belonging to one of the following categories:

skeleton that reflects no more than the basic form of the pattern, but if no special precautions are taken the skeleton will usually have some odd strokes or 'tails' that have nothing to do with the basic shape. *Fig. 4* gives an example of such a skeleton; it contains many end points and fork points at illogical places, and therefore recognition from the special points will obviously not be possible in this case.

The shape of the skeleton also depends, of course, on the initial pattern, and thus on the way in which it is formed from the television picture. Since we only have occupied and unoccupied positions in the matrix, the grey tints will be lost from the television picture; above a certain threshold value they will be regarded as black, and below it as white. The choice of this threshold value may affect the shape of the skeleton produced from the pattern. In our investigation this threshold was given a fixed value, but a better result would be expected if the threshold value were allowed to adjust itself automatically to the blackness distribution in the television picture [9].

We have successfully developed a thinning procedure that produces very useful skeletons (see *fig. 5a, b* and *c*). Even these skeletons may sometimes have a 'tail', but this always corresponds to a distinct bulging of the unthinned numeral, as can be seen in the top right-hand corner of fig. 5c. The details of our thinning procedure will be given in part II of this article, which will also give a description of the method of determining the special points of the skeleton.
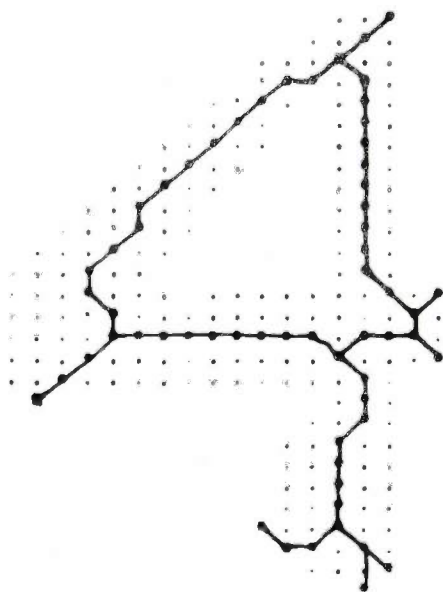
Fig. 4. Example of a skeleton that cannot be used for the recognition procedure because it has a number of 'tails' that have nothing to do with the basic shape of the figure to be thinned. The picture shown here was made by the line printer of the computer used for the skeletonization process. The dots indicate the points of the original figure that were removed during the skeletonization. After the print-out, the skeleton points were joined up by hand to make the skeleton visible.

Since there is not space here to describe a complete recognition procedure, we must confine ourselves to a simple procedure capable of recognizing a limited assortment of numeral types. It will then be shown how a procedure of this kind can be used as a starting point for the development of a procedure for recognizing numerals written with no restrictions.

A simple recognition procedure

To recognize numerals of the types shown in *fig. 6a* it is sufficient to use the simple recognition procedure of fig. 6b. In this the numerals are first divided into groups corresponding to the number of end points found in the skeleton — this number will be denoted by E — and to the number of fork points — denoted by the letter F. If such a group comprises different
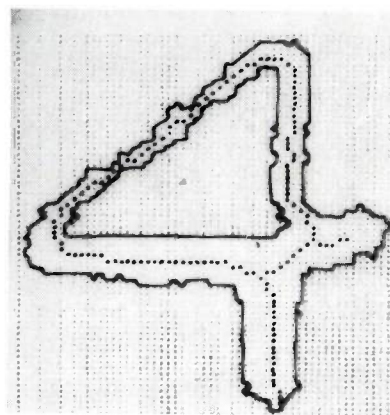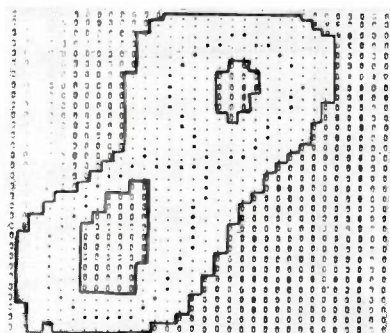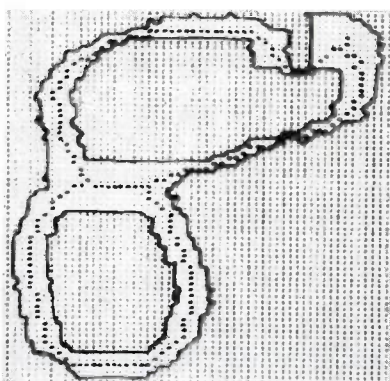
Fig. 5. Three examples of skeletons that are suitable for our recognition procedure. The figures were traced out in the same way as in fig. 4 by the line printer, but in addition the unoccupied positions of the matrix have now been marked with an 0. After printing, the skeleton points were marked by hand and the edge of the original figure was drawn in. Crossings (four-junctions) in the original numeral usually degenerate into two three-junctions (fork points) (*a*); if the skeleton does have a four-junction (*b*) it is counted as two fork points. The skeleton of (*c*) has a tail at the top on the right that corresponds to a marked bulge in the pattern.

[8]   See the first article of note [4].
[9]   See the first article of note [2].

a

0   1   2   3   4   5   6   7   8   9

2   3   4   5   7   8

Fig. 6. *a*) Collection of carefully written numerals, for some of which two widely used types are included. This is of course only a very small selection from the many possible ways of writing these numerals. *b*) Procedure for recognizing numerals of the types in (*a*). The numerals are first classified by the number of end points (*E*) and the number of fork points (*F*). If a numeral belongs to an (*E*,*F*) group that includes more than one numeral, a subsidiary recognition procedure is used to distinguish between these numerals.

(E,F)-GROUP	TYPE OF NUMERAL	PROCEDURE	RECOGNIZED AS

(0,0)   0   0

(0,2)   8 8   8

(1,1)   6 9   [Is e above f ?] — yes → 6 / no → 9

(2,0)   1 2 3 5 7   [Is W equal to H ?] — yes → 1 / no:
2357 → [Is e₁ on the right-hand edge of the frame ?] — yes → 5 / no:
237 → [Is e₂ on the right-hand edge of the frame ?] — yes → 2 / no:
37 → [Is e₂ on the lower edge of the frame ?] — yes → 7 / no → 3

(3,1)   2 4 5 3 4 7   [Is f in the upper third of the right-hand half of the rectangle ?] — yes → 7 / no:
23445 → [Is f in the lower third of the rectangle ?] — yes → 2 / no:
3445 → [Is f to the right of the three end points ?] — yes → 3 / no:
445 → [Are e₁ and e₂ above f ?] — yes → 4 / no:
5 → [Is e₁ on the right-hand edge of the frame ?] — yes → 5 / no → rejected

b

numerals a subprocedure is used for further selection. Fork points as used here refer exclusively to three-junctions; a crossing in the original numeral in the form of a four-junction usually degenerates in the skeleton into two fork points (see fig. 5a). On the rare occasions when a skeleton contains a real four-junction (see fig. 5b), we count it as two fork points.

In addition to E and F there is a third and equally obvious characteristic of numerals; this is the number of closed loops. We do not use this characteristic in our method, but that does not mean that we neglect any essential information, for there is a simple relation between the quantities E and F, the number of loops L and the number of separate sections S in the figure. This relation, which is easy to verify, is:

$$F - E = 2(L - S).$$

Normally the pattern (the numeral) is in one piece, so that $S = 1$. Once we have determined F and E, the difference $L-S$ is then a fixed quantity, but not L and S themselves. This is easily understood, for if a closed loop is added as a new section, F and E do not change, but L and S are both increased by 1.

We shall now deal in turn with the various groups (see fig. 6b).

$E = 0, F = 0$ (group 0,0).

Only a *nought* can possibly give a skeleton of this kind. The *noughts* are thus recognized (since here we are only concerned with the carefully written noughts of fig. 6a).

$E = 0, F = 2$ (group 0,2).

This applies only to the *eights*, which are thus recognized with this group.

$E = 1, F = 1$ (group 1,1).

This applies to the *sixes* and the *nines*. A simple and sufficient criterion for making the correct choice is found in the relative positions of the end point and the fork. If the end point e is higher than the fork f, we then have a *six*, otherwise we have a *nine*.

$E = 2, F = 0$ (group 2,0).

This group includes the *one*, and also the *two*, the *three*, the *five* and the *seven* of the *top* row in fig. 6a, none of which have a fork. To distinguish between these numerals we first count the total number of points (squares) forming the skeleton. We call this number W (for weight). Next we count the number of rows that the skeleton covers in the matrix, and we call this number H (for height). For all *ones* we have: $W = H$, but this does not apply to the other numerals. We have thus eliminated the *ones*.

To select from the other possibilities, we number the end points in the sequence in which they are en-

countered when the matrix is scanned row by row, beginning with the top row and moving from left to right. The end points are given the symbols e_1 and e_2. We then enclose the skeleton in a rectangle formed from the horizontal and vertical lines that touch the skeleton. If e_1 is at the right-hand edge of this frame, the numeral must evidently be the *five* from the top row of fig. 6a; if, however, e_2 falls here, then the numeral is a *two*. If neither e_1 nor e_2 are at the right-hand edge of the frame, the decision is between the *three* and the *seven*. Because of the limitation to the types of numeral given in fig. 6a, we need in this case only look to see whether e_2 is at the bottom edge of the frame. If this is so we then have a *seven*, otherwise it is a *three*.

$E = 3, F = 1$ (group 3,1).

This group includes the two *fours*, and also the *two*, the *three*, the *five* and the *seven* of the *lower* row in fig. 6a. We look first to see whether f (the fork) lies in the right-hand half of the upper third of the rectangle. If it does, then we have a *seven*. Next we see whether f is in the lower third of the rectangle. In this way we identify the *twos*. We then see whether the three end points are all on the left of the fork, which enables us to recognize the *threes*. We find the *fours* by seeing whether e_1 and e_2 both lie *above* f. Any numeral that is still not classified can only be a *five*. We have included a further safety criterion here, however, to make sure that e_1 does in fact lie on the right-hand edge before a *five* is decided upon. If it does not, the numeral is rejected as unrecognizable.

In the procedure of fig. 6b all numerals are rejected that have different numbers of end points and forks than those that correspond to one of the five combinations above. If the only numerals presented are of the types included in fig. 6a as in the original requirement, no numeral will be rejected, and the safety criterion is then superfluous. Our purpose in adding it was simply to introduce the concept of 'safety criterion', since in general a procedure for recognizing numerals written without restrictions can contain a large number of such safety criteria.

If the recognition procedure described here is applied to an *arbitrary* set of numerals, the result will be a number of correct readings, a number of errors (substitutions) and a number of rejects. All the numbers in the first block in fig. 1 were correctly recognized, but these 70 numerals were written with the procedure of fig. 6 in mind. It does however show that the automatic recognition of handwritten numerals would be a fairly simple matter if the people that write them could be persuaded to observe a few simple rules. We shall return to this point later on in the article.

The success of our method is largely due to the subdivision into groups before the actual recognition begins. The only characteristics used for this classification are the number of end points and forks. The recognition of arbitrarily written numerals requires a complicated scheme for each of these groups, to be able to distinguish the many numerals belonging to them. We could increase the number of groups by using various other appropriate characteristics of numerals in the classification, such as the presence of a 'tail', etc. The groups would then be smaller and the subprocedures propably simpler. Perhaps a better result might be obtainable in this way for the whole system than with a classification based only on special points. The possibilities of extending the number of groups in such a manner have not yet been fully examined.

Refining a recognition procedure

In developing a recognition procedure for numerals written without restrictions we start from a simple procedure like the one described in the previous section. This is applied to a number of numerals, some of which will be read wrongly or rejected. We now devise some new criteria to enable these numerals to be properly recognized as well, and add these criteria to the procedure. We then apply this modified procedure to a new group of numerals, and again introduce modifications that will permit all these new numerals to be recognized. Since this process is repeated a few times the procedure will gradually improve, but at the same time it will become more complicated. In making adjustments we will undoubtedly devise criteria that could more usefully have been added to the procedure earlier. When it becomes obvious that the whole procedure is becoming impossibly complicated, the time has come to throw it into the wastepaper basket and then, armed with the experience gained, to start on a new one. The new procedure will be more effective than the old one, and at first simpler, but after some time this procedure will also have to be discarded to make way for yet another one. This discarding is one of the most important but also one of the most difficult processes in the development of a recognition procedure.

The adjustment process will be explained with a single example, showing at the same time that the use of the skeleton by no means implies a limitation to the end points and the forks, but that in fact other special points can also usefully be employed for recognition. We consider the *nine* and the *nought* from *fig. 7*. On attempting to recognize these numerals with the aid of the simple procedure described in the previous section, we see that they both belong to the group (1,1) and both are classified as *nine*, which is of course unacceptable. We now scan the matrix row by row, beginning with the row on which *e* lies and then moving up to the row on which *f* lies. We look for the first row on which only one skeleton point lies, and once we have found
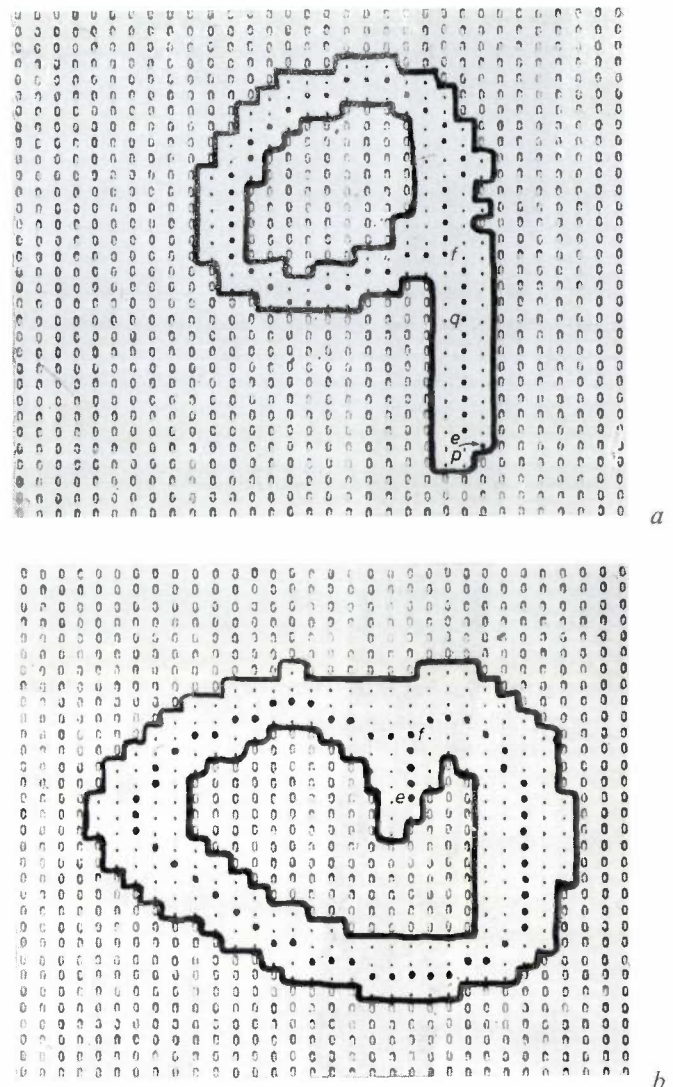
Fig. 7. Example of two numerals that are not distinguished by a simple recognition procedure like that in fig. 6*b*. The numeral in (*a*) is correctly recognized as a *nine*. The numeral in (*b*) is meant to be a *nought*; here, however, a bulge in the loop has caused a tail. Since the end point of this tail lies lower than the fork, this numeral is also classified as a *nine*. To avoid this error we add a new criterion to the recognition procedure to investigate whether the tail is *in* or *below* the loop.

that point we mark it as the special point *p*. In fig. 7*a* the point *p* coincides with *e*, in fig. 7*b* we find no point *p*. Next we scan the matrix again, now moving downwards row by row beginning with the row below *f*, and we go no further than the row with *e*. Here again we look for the first row with only one skeleton point; if such a point exists we call it *q*. We find such a point *q* in fig. 7*a*, but not in fig. 7*b*. In the further process of recognition we can now make use of the existence or non-existence of the points *p* and *q*, and of their positions. To solve the problem of fig. 7 it is sufficient to ask whether a point *p* exists. If it does, we have a *nine* (the numeral then has a tail projecting downwards), otherwise we have a *nought*. With this extension the

recognition procedure is still far from perfect, but it is slightly improved.

To recognize a numeral we have so far confined ourselves to the characteristics of the skeleton, but this is not possible in all cases. It regularly happens that the typical features of the original numeral get partly or completely lost in the skeleton. Let us take one example. *Fig. 8* shows a *nine* whose skeleton, owing to a slight discontinuity on the right-hand side at the top in the 'thick' numeral, has the same basic shape as a *five*. A choice between these two numerals, purely on the grounds of the skeleton, is very difficult in this case; we must therefore use the thick numeral. We could for example add a criterion that enables us to decide, from the size of the break in the loop in the thick numeral, whether it is a *five* or a *nine*.

An unfortunate aspect sometimes encountered in the development of a recognition procedure is that an improvement in one point makes another point worse. It is therefore dangerous to let the procedure include a type of numeral that occurs infrequently if it at all resembles a frequently used type of another numeral (inclusion of the *nine* from fig. 8 endangers the recognition of the *five*). It is best to delay this until very many numerals (e.g. hundreds of thousands), of widely different origin have been tested.

Provisional results

In total, more than 25 000 written numerals have been used in our investigations. Eighty to ninety per cent of these numerals present no problems with our method. The remainder show a wide variety of less conventional shapes, some of which were encountered very infrequently, others more often, although they were then usually from the same person. From the limited number of numerals available to us we were not able to conclude anything about the frequencies with which these aberrant forms of numeral will occur in practice. Consequently it is not possible to say exactly what percentage of correctly recognized numerals will be obtainable with our method. Nevertheless some general conclusions can be drawn about the practical usefulness of the method.

During our investigations various recognition schemes were developed. The most extensive scheme now available at Philips Research Laboratories for the recognition of arbitrary numerals was designed from more than 15 000 handwritten numerals. Most of these were placed at our disposal by the Dr Neher Laboratory of the Netherlands Post Office; the remainder came from forms that were filled in by colleagues and by visitors to demonstrations. The recognition scheme was checked against 10 000 numerals that were *not*
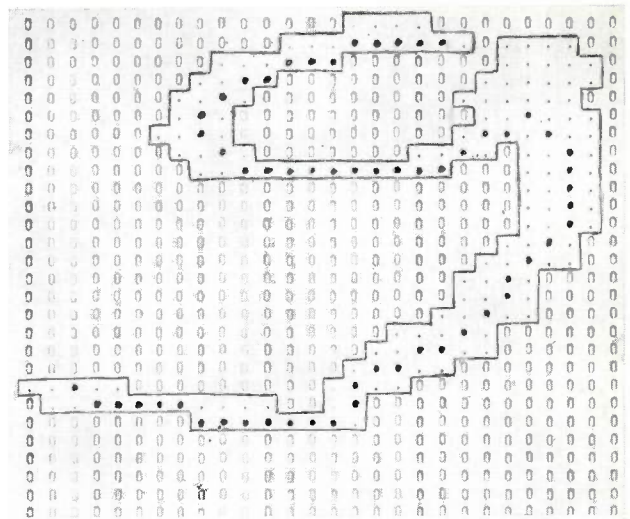
Fig. 8. Example of a numeral where the skeleton is not sufficient for correct recognition. Because of the chance break in the loop, the skeleton of this *nine* has the same basic form as that of a *five*. To classify this numeral properly we must add a criterion that relates to the 'thick' numeral, e.g. by determining whether the break in the loop is very small.

involved in the development of the scheme. These numerals also came from the Dr Neher Laboratory. This test material was obtained in a door-to-door survey, in which people were asked to fill in a particular series of numerals in spaces on a printed form. With this method it was known what numeral the writer intended each time. No restrictions were imposed on the handwriting, and since the people filling in the forms had no personal stake in the results, the average quality of the handwriting was fairly poor (see *fig. 9*). The result of the test is presented in *Table I*; it can be seen that 91.37% of the numerals were recognized, 2.67% were read wrongly (confused with another numeral) and 5.96% were rejected as unrecognizable.

Table I. Results of automatic reading of 10 000 numerals. Each space shows how many numerals which the writer meant to be read as the numeral on the left were read as the numeral above. The last column contains the numbers of rejected numerals; the numbers of numerals read correctly appear along the diagonal.

	0	1	2	3	4	5	6	7	8	9	*
0	1253	1		1		1			9		25
1		1042	3		1	3		25			29
2		1	1123	6				6	1	7	101
3	1		5	802		14		5	1	11	84
4	1	2	1		624	3		11		14	57
5				2		1214			1	4	58
6	9	2	1			5	662		3		31
7		20	9	7	9	4		765		12	94
8	6	1	1	1	1			7	1013	5	64
9		2	1	11	1	2		5	1	639	53

correctly read   :   9137 numerals (91.37%)
erroneously read: 267 numerals (2.67%)
rejected   :   596 numerals (5.96%)
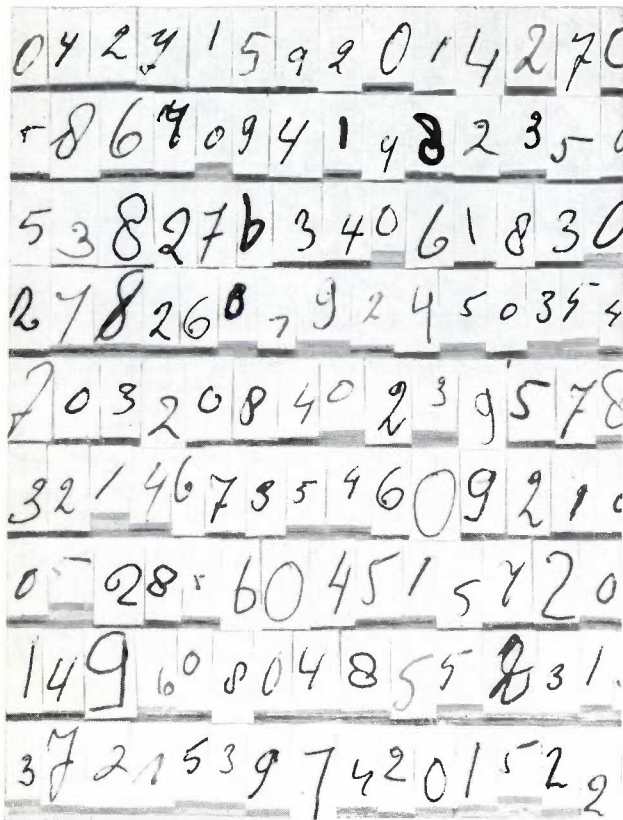
Fig. 9. Example of the test numerals obtained from a door-to-door survey. The numerals had to be written in spaces on printed forms. To give an idea of the quality of the handwriting, the figure shows a number of numerals taken at random from different forms (forms of three types were used in the survey, with spaces of different sizes).

The origin of the numerals has been described at some length because in judging the performance of a machine for recognizing handwriting it is of the utmost importance to know exactly what test material was used. This seems obvious but we have found by personal experience how easy it is to deceive oneself and others in this kind of experiment. In fact, a character-recognition system can only be adequately judged when it has been used in practice for millions of numerals.

Evaluation of the results

A result of not quite 92% of numerals recognized is of course far from sufficient for practical purposes. By including large numbers of new numerals in our investigations we would certainly be able to increase the percentage quite considerably, but the effort this costs increases with the improvement achieved. In our opinion it will never be possible to make the step from 99 to 100%. To see whether there is any point in seeking further improvement, we have tried to determine what a particular percentage of correctly recognized numerals means in practice.

Reading forms

In general, the forms that have to be read have not one but several numerals written on them. The requirement then is of course that all the numerals are recognized. If however the individual numerals are not recognized with 100% certainty, then the percentage of forms read correctly will decrease rapidly as the number of numerals written on them increases. For example, if the percentage of numerals recognized is 98, then 90% of forms with 5 numerals will be correctly read, but only 78% of forms with 12 numerals. To read correctly 99% of forms with 12 numerals (the average on Dutch giro cards) it will be necessary to recognize no less than 99.9% of the individual numerals. This seems an impossible task; even a human reader will not be able to reach this percentage if he has no means of making any comparison between the numerals written on a form — which the machine does not do.

The various percentages are calculated in the following way. A form on which a number of numerals are written is only correctly read if all the numerals are correctly recognized. It is rejected if one or more numerals are rejected and it is wrongly read if one or more numerals are wrongly read and no numeral is rejected. If we take the proportion of correctly read numerals as c_n, the proportion rejected as r_n, and the proportion read in error as e_n, we can find the corresponding proportions c_f, r_f and e_f for forms with m numerals from:

$$c_f = c_n{}^m,$$

$$r_f = 1 - (1 - r_n)^m$$

and

$$e_f = 1 - (c_f + r_f).$$

Fig. 10 gives a plot of the curve

$$y = 1 - (1 - x)^m$$

for a number of values of m. Given c_n and r_n we can easily read the proportions c_f, e_f and r_f from this graph.

In making these calculations we assumed that the numerals that were difficult to recognize were randomly distributed over
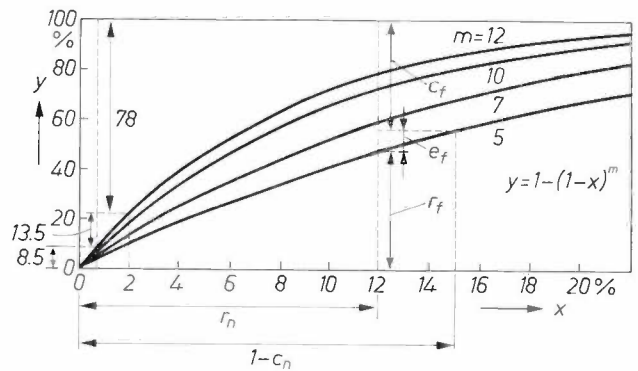
Fig. 10. Illustrating the derivation of the proportion of correctly read forms c_f, those read wrongly e_f and those rejected r_f, from the corresponding proportion c_n, e_n and r_n for individual numerals, on forms with m numerals. The proportions r_n and $(1 - c_n)$ are plotted horizontally; the desired proportions can be read from the graph in the way indicated. The numerical example given refers to the result of the reading test described in fig. 11.

the forms. It is not unreasonable to expect, however, that in practice the rejected and wrongly read numerals will tend to be concentrated on cards filled in by people who write carelessly. This would make the number of wrongly read forms much lower than follows from the calculation. In fact, however, it turned out that this was not the case at all — at any rate with the procedure used for the test.

A human reading test

To get some idea of the accuracy with which people can read the numerals used in our investigations, 35 people were subjected to a reading test. In this test 840 numerals were used taken at random from the group of 10 000 numerals with which we tested our recognition method. Some of these 840 numerals are shown in fig. 9. The figure indicates the method of presentation; the subjects could not see which numerals had originally appeared together on the same form, so that they were not able to recognize numerals by comparing them with other numerals in the same handwriting — a possibility that does not occur in our

the numerical example in fig. 10). This does not of course mean that people are unable to read giro forms; on average the numerals will be written more clearly than our test numerals were, but the most important point is that people can compare the numerals on a form.

Provisional conclusions

Without wishing to draw any final conclusions from the above, it does seem that the results clearly indicate that the problem of reading giro cards entirely by machine — the problem always in mind during the investigation — cannot be solved simply by further improvement of the recognition procedure for individual numerals. It is hardly likely that the performance of the machine will ever be better than the average result of our 35 test subjects, and even this result gives little reason for optimism. This does not imply, however, that the method in its present form is useless. In the giro service in the Netherlands all cards are pro-
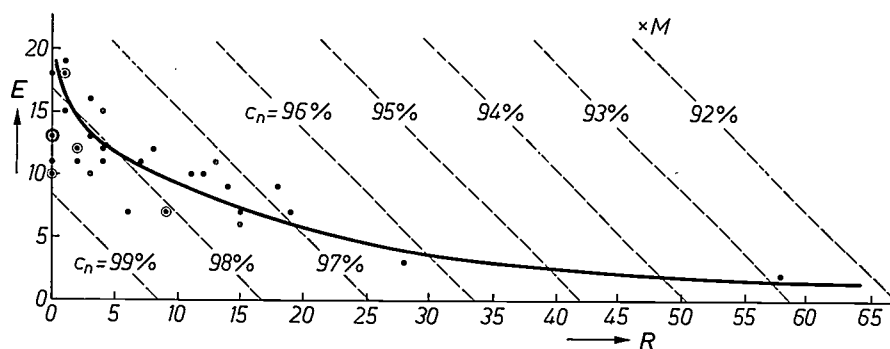
Fig. 11. Result of a reading test in which 840 numerals were read by 35 people. The numerals were taken at random from the material used to test our method (see fig. 9). The number of rejected numerals R is plotted along the horizontal axis, and the number of numerals read wrongly E is plotted vertically. The lines at 45° give the proportion of correctly read numerals c_n. Each point gives the result for one person; a circle around a point indicates that two people scored the same result. An approximate relationship between the number of rejects and errors can be obtained from the points (see the solid curve). For comparison the figure also shows the result given by our recognition procedure with these 840 numerals (point M.)

recognition method either. *Fig. 11* gives the result of the test; on average 98% of the numerals were correctly read, 1.3% were read wrongly and 0.7% were rejected. The same 840 numerals were read by the machine (using the same recognition scheme as tested above) and the percentages found were 91.7, 2.6 and 5.7 (point M in fig. 11).

The result of the reading test shows that the machine does not as yet rival the capability of a human reader. In comparing man and machine, however, we should bear in mind that the results of the reading test are somewhat biased by the fact that in doubtful cases a human reader can guess, which the machine does not do. With the percentages found in the reading test with human readers, only 78% of forms with 12 numerals are read correctly, 13.5% wrongly and 8.5% are rejected (see

cessed independently of one another by two punch operators, and their results are automatically compared. One of the punch operators might perhaps be replaced by a machine using the most universal recognition procedure available. The results of the machine can be further improved by on-line correction, with the rejected numerals displayed on a monitor so that a punch operator can read them. This would only reduce the number of *rejected* numerals of course, not the number of errors.

It might be interesting to consider extending the system to include new facilities to increase its basic capabilities. If a memory function were built into the system the machine could compare the different numerals written on a form, like a human reader. This would perhaps allow a decision between doubtful

alternatives, which may occur with numerals like 1 and 7. All this, however, is still rather speculative. As far as we can see the situation at present, there is no likelihood as yet of faultless automatic recognition of numerals written without restriction, and a little cooperation from the public will be needed to obtain complete automation.

Where an automatic reader is to be used exclusively within a closed organization, for example only for the administration departments of a firm, the handwriting can then be subjected to constraints designed to ensure that the automatic reader will recognize all numerals correctly. This can be done by choosing types of numerals that are suited to the writing habits of the people in the department, and programming the machine with a recognition procedure that recognizes all these types of numerals. If the reader now rejects numerals, or makes errors, it means that someone has not kept to the rules. The machine has then been adapted to the users. One can of course adopt the opposite reasoning and design an automatic reader that recognizes only a limited number of types of numeral. The users must adapt themselves to the reader but the recognition procedure can be made so obvious that the users can easily understand it and can easily learn how to write the numerals.

At Mullard Research Laboratories, which are affiliated with us, a simple recognition procedure based on our ideas has been developed for use in a character-recognition system. This system (the Philips X 1300), which is primarily intended for reading standard (OCR) characters, is being manufactured by the M.E.L. Equipment Co., Ltd., a part of the British Philips Group.

A system imposing constraints on the freedom of writing could also be devised for giro services. The constraints imposed in such a case, however, would have to be kept to the minimum. The principal requirement would be that forms on which the writing did not follow the rules would have to be rejected by the automatic readers. This could perhaps be arranged by providing the giro card with a special space in which the account holder makes a mark to show that he has observed the rules and agrees to automatic reading. All the cards are then read automatically, and the automatically read numerals from the *marked cards only* are used in the giro records. The numerals from the unmarked cards are also read by human readers and only their results are used in the giro processes. However, the automatically read numerals are given on the account holder's statement. In this way the account holder has a chance of becoming familiar with automatic reading before consenting to the use of the system. In return for the care required from the account holder he could be offered some recompense, such as faster handling of his payment orders.

A high degree of certainty that a form on which the writing does not follow the rules will be rejected can be obtained by combining the recognition system with a number-entry system

Fig. 12. Strip with spaces for filling in numerals by the Van der Toorn system. Each space is marked with the two numerals that may be entered in it. The number 2503 has been filled in as an example; the 2 is in the first space that may contain a 2, the 5 in the next space along that may contain a 5, and so on. When reading a numeral in a particular space the recognition procedure therefore no longer has to make a choice from ten numerals, but only from the two that are allowed to be entered in the relevant space. This considerably reduces the number of errors and also — if the recognition procedure has been adapted to this numeral-entry method — the number of rejects.

proposed by L. van der Toorn [10]. To illustrate this system let us take as an example a case in which the number 2503 is to be filled in. A strip of spaces like that shown in *fig. 12* is used; first we look for the space containing the numerals 2 and 3. These numerals indicate that a 2 may be entered in this space, or a 3, but no other numeral. The first digit of our number is therefore written down in this space. Next we look along the strip — going to the right — to find the first space in which we can place the 5 of our number and so on. While the numbers remain easily readable for people, the task of the machine is lightened enormously, since it now no longer has to decide from *ten* choices but only *two*. The location of the space indicates which pair it is. We can now use a recognition procedure that is simpler and more effective; doubtful cases such as 1 and 7, which increase the number of rejects in a normal scheme, do not now arise.

As an illustration we shall first combine Van der Toorn's entry system with the recognition scheme which we tested above. With the 10 000 figures used for this test the percentage of wrongly read numerals falls from 2.67% to 0.21%. This may be counted in Table I, where we now take only the confusions of 0 and 1, 2 and 3, etc. as errors, and all the others as rejects. The number of errors is indeed very small, but there are now more rejects (8.42%). *Table II* gives the result for a recognition procedure adapted to the number-entry method. Although it was still in a very primitive state of development it gave much better results. There are now 0.75% errors and only 1.76% rejects.

The system works even better if each space is marked with one

Table II. Results of automatic reading of the same 10 000 numerals as used for Table I, but now for the combination of our method with Van der Toorn's method of filling in numerals. The recognition procedure used was specially designed to distinguish between the two numerals that may be entered in the same space.

	0	1	2	3	4	5	6	7	8	9	*
0	1273	5									10
1	1	1095									6
2			1221	6							16
3			6	904							13
4					662	9					42
5					10	1249					21
6							696	2			17
7							9	891			22
8									1073	7	19
9									20	685	10

correctly read   :   9749 numerals (97.49%)
erroneously read: 75 numerals (0.75%)
rejected   :   176 numerals (1.76%)

[10] Netherlands patent application No. 6918149.

numeral instead of two. When the machine reads a numeral, then the locations alone of the space containing the numeral gives unambiguous identification of the numeral, so that the recognition is given the nature of a verification. This is a combination of two entirely different methods of recognition; a method that classifies by location (in the same way as in a mark-sensing system), and a method that classifies by shape.

A serious difficulty with this version of the number-entry system is that twice as many spaces are needed for the same number of digits. Even when each space is marked with two numerals, it is difficult enough to find room for the strip on forms in current use; using only one numeral per space it would be impossible.

It can easily be seen that Van der Toorn's system meets the requirement that the machine will reject forms filled in wrongly. If no notice is taken of the rules for filling in the numerals, or they are not understood, the chance is extremely small that one of the two permitted numerals will be read in each of the squares. If more numerals than we have available at the moment could be used for making the recognition scheme, it would be possible to reduce the percentages of errors and rejects very considerably, and the system could then become a really practical proposition.

Summary. A method is described for the automatic recognition of freely handwritten numerals. The numerals are scanned one by one with a television camera, and the television picture is then electronically converted into a matrix of occupied and unoccupied positions. This matrix is very suitable for processing in a computer or in electronic equipment specially designed for the purpose. The numeral is first 'thinned' to a skeleton that is nowhere thicker than one matrix position. The end points and forks in the skeleton are then determined. These special points form the basis of the recognition procedure that is used to classify the numeral, and is based on a large number of test numerals. In certain cases the skeleton alone is not sufficient for recognition, in which case tests must be made on the unthinned numeral. The system described has great flexibility since the recognition procedure can continuously be elaborated upon by the addition of new criteria. A recognition procedure drawn up from more than 15 000 numerals was checked against 10 000 other numerals; 91.37% of these were recognized, 2.67 were wrongly read and 5.96% were rejected. A reading test is described in which 35 people had to read a number of numerals taken at random from the test group of 10 000. Without being able to compare numerals written in the same handwriting, they were able to read 98% of the numerals correctly. Since this is not sufficient for reading forms with say 12 numerals, like giro cards, and since it is not to be expected that a character-recognition system will perform better than a human reader, it is concluded that if forms are to be read entirely by such a system, either the method must be extended, e.g. by building in a memory function for comparing numerals, or constraints must be imposed on the freedom with which the numerals are written. Some possibilities of devising a practical system are indicated, both for a closed organization and for the general public.

Improved ear-lobe clip for physiological transducers

In medicine and for ergometric measurements transducers are sometimes mounted in a small clip so that they can be connected to a part of the body such as an ear lobe [1]. The transducer usually consists of a light source and a detector, and the amount of light transmitted through the ear lobe is measured. Since the light transmitted depends on the amount of blood in the ear lobe, such a transducer is suitable for observation of the heart rate.

In practice an appreciable fraction of the signals picked up by the transducer are due to unwanted effects, caused by movements of the subject, that make the clip move with respect to the ear. If more information than the heart frequency is to be determined from the signals, the unwanted signals ('artefacts') should be kept as low as possible. This was the starting point in the work on the new ear-lobe clip described here.

Movement artefacts can be largely suppressed by making the pressure of the clip very high, but this is not very comfortable for the subject. We have produced a different kind of solution to the problem (*fig. 1*). The light source and the detector are not mounted in the actual 'fingers' of the clip, but in a 'cup' of plastic shaped rather like a hemisphere. Each cup is supported at three points located as far apart as possible, and is fastened to the finger of the clip by a steel wire that passes through a hole in the cup. The clearance between wire and hole is made large so that when the subject puts the clip on, the cups can take up a position with the flat part pressed well against the ear. Once the clip has been put on this position remains fixed because of the friction between the cup and its support points; any extra pressure on the cup due to movement of the subject also increases the frictional forces and its position is unaffected.

The second special feature of the new clip is that the friction in the hinge can be preset by the user. This means that the clip can be fitted with a spring that is stronger than in other clips, yet nevertheless does not compress the ear lobe any harder. It ensures that even relatively large forces, which the ear lobe might exercise on the clip, will not make it open further — the second source of artefacts.

To fix the position of the cups the force exerted on them by the ear lobe should not be less than 0.1 N, while a value of more than 1 N is uncomfortable for the subject. Thus, if the clip force is made about 0.5 N and the frictional force about 0.3 N, the subject feels only 0.2 N, which is not uncomfortable. External forces up to 0.8 N cannot however open the clip.
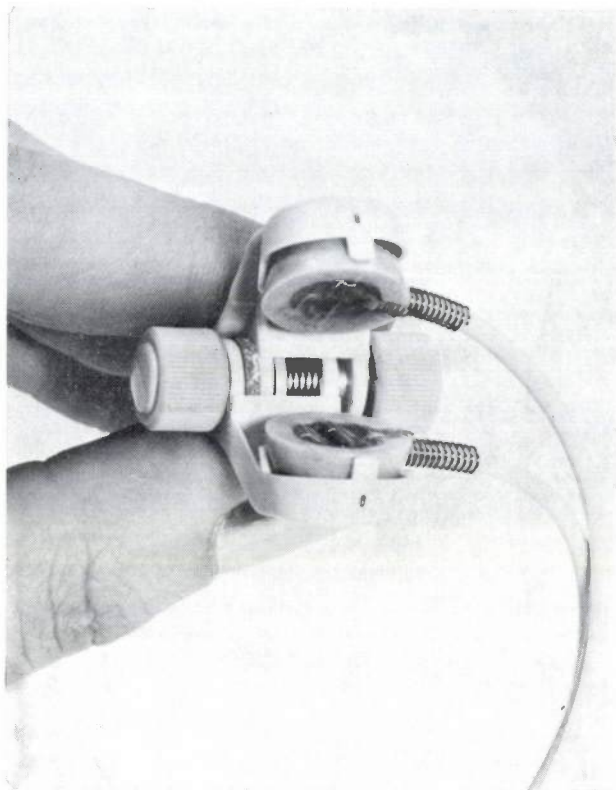
Fig. 1. *a*) Exploded view of the new ear-lobe clip for a physiological transducer, which allows a measurement signal practically free of artefacts to be obtained. The light source and the detector, which form the transducer proper, can be mounted in the cavities in the 'cups' shown in the middle of the drawing. Each cup is fastened to its finger of the clip by a thin steel wire that passes through three holes, two in the finger and one hole with plenty of clearance in a projection on the cup. With this very free arrangement the flat side of the cup can readily take up a maximum-contact position against the ear lobe when the subject puts the clip on. From then on this position is maintained by the friction between the cup and its three support points, even if external forces act on the clip. *b*) The part of the cup that must contain the point of action of an external force if the position is to remain fixed. This region extends almost to the edge near the three support points. Forces with points of application elsewhere can alter the setting of the cup if they are large enough to overcome the friction. *c*) Cross-section showing central pin, spring (*S*) and the friction adjustment. The pressure of the fingers of the clip on the friction washer (grey) can be varied by the two knobs
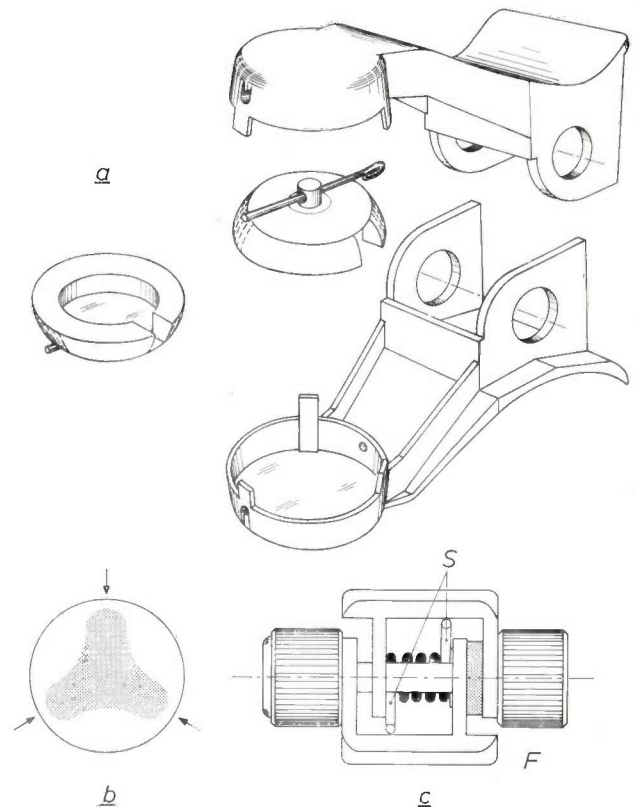
The forces that operate on the clip when the subject moves his head are kept as small as possible by making the clip of extremely light material — a reinforced polycarbonate plastic — and by making the leads very light and flexible. The clip with its light source and detector weighs less than 5 grams.

The light source can be a light-emitting GaAs diode (an 'LED') mounted in one of the cups; this diode requires only 37.5 mW. The silicon photodiode BPX-41 is a suitable detector; its sensitivity outside the infrared is very low [2]. If the LED is pulsed-operated, synchronous detection can be used; this then further reduces the effect of daylight on the detector. The transducer with these two elements is very suitable for use with the Philips Modular Patient-Monitoring System.

The pen-recorder traces in *fig. 2* show clearly the improvement to be obtained with the new clip; very few artefacts appear even when the subject shakes his head hard.

<div align="center">C. P. van Nie</div>
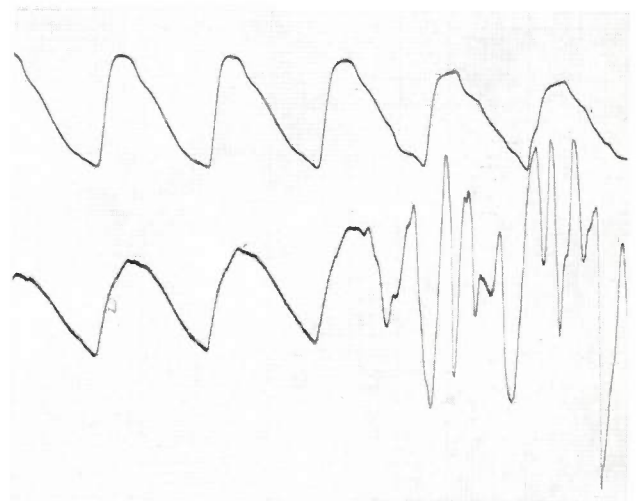
Fig. 2. Part of a recorder trace of the pulse signal for the same subject simultaneously registered by a conventional transducer and the new one. During the second half of the recording the subject shook his head hard. In one curve the heartbeat can hardly be distinguished; in the other the artefacts are barely visible; if the subject only makes normal movements the artefacts are an order of magnitude smaller. With the new transducer the magnitude and waveform can be determined reliably from the trace, as well as the frequency.

Ir C. P. van Nie is with the Philips Medical Systems Product Division, Eindhoven.

[1] See for example: G. A. Harten and A. K. Koroncai, A transistor cardiotachometer for continuous measurements on working persons, Philips tech. Rev. **21**, 304-308, 1959/60.
[2] The idea of using an LED with a photodiode for this kind of measurement came from Ir M. van Herk of Philips Research Laboratories, Eindhoven.

104

Philips tech. Rev. 33, 104-113, 1973, No. 4

Heat pipes

G. A. A. Asselman and D. B. Green

I. Operation and characteristics

II. Applications

Heat-transport problems in nuclear reactors and in space applications have in recent years stimulated intensive research in many countries on a heat-transport method whose principle has long been known, the method of transporting heat in its latent form, i.e. the latent heat of vaporization. Heat conductors based on this principle, called 'heat pipes', can have a thermal resistance more than 10 000 times smaller than that of a copper rod of the same dimensions. Heat pipes have been studied at Philips Research Laboratories and are of considerable interest for heat transport for Stirling engines and the development of isothermal ovens.

Although it has long been known that high rates of heat transport can be obtained by means of the evaporation-condensation process, it was not until 1944 that the first patent appeared on a 'heat pipe', that is to say an evaporation-condensation device with recycling of the fluid by capillary action [1]. The first paper on the heat pipe came only in 1964 [2]. Since then, interest in this field has grown rapidly. All manner of applications have been envisaged, probably more than for any device with the exception of the laser.

Unfortunately many of these potential applications will only be feasible with a working fluid possessing the density of liquid hydrogen, the latent heat of lead and the viscosity of superfluid helium. In the meantime, however, heat pipes have been made that are based on materials that do in fact exist. These materials and the principles of operation of a heat pipe are discussed in this article. Part I is primarily concerned with the theory, while part II will describe a number of applications [*].

I. Operation and characteristics

The heat pipe is a device whose primary function is the transport and distribution of heat by means of the evaporation and condensation of a working fluid. Its most characteristic feature is that it derives the power necessary for the circulation of the liquid and vapour in the presence of gravitational and frictional losses, solely from the heat input; hence no external pumping system is required. In this respect the heat pipe ranks in terms of simplicity with the domestic coffee percolator. However, unlike a percolator, which accomplishes the circulation of coffee by means of a bubble pump and is thus dependent upon a gravitational field for its action, the heat pipe employs a capillary structure, or 'wick', and is thus a device of far greater flexibility in both design and application.

In practice the heat pipe is usually no more than a tube or duct whose wall, or a large part of it, is clad with a layer of porous material, the wick just referred to (*fig. 1*). The duct is heated at one end, causing the liquid to evaporate. This end is thus called the 'evaporator' *E*. The other end is the 'condenser' *C*, which is connected with the device or component to which the heat is to be transported. The vapour condenses at this end, and gives up its latent heat. In an operating heat pipe there is a continuous flow of vapour through the duct from evaporator to condenser, and liquid is continuously recycled in the opposite direction back to the evaporator via the wick. The driving force for

[*] Part II of this article will appear in the next issue of this journal.
[1] R. S. Gaugler, U.S. Patent 2350348, 8 Jan. 1944.
[2] G. M. Grover, T. P. Cotter and G. F. Erickson, J. appl. Phys. 35, 1990, 1964.

Ir G. A. A. Asselman is with Philips Research Laboratories, Eindhoven; Dr D. B. Green was formerly with Philips Research Laboratories.

this liquid flow is the capillary action of the channels in the wick, and therefore it is essential that the working fluid should wet the wick. The working fluid can be anything from liquid hydrogen to liquid silver, depending on the temperature range within which the heat pipe is required to operate.

Since both evaporation and condensation can occur at almost the same temperature, the total temperature drop over a heat pipe may be very small, a few degrees or so, whereas the heat transported may be of the order of kilowatts. The effective thermal conductance can, therefore, be very high — as much as thousands or tens of thousands of times that of a copper rod of the same size.

It is perhaps useful to point out at this stage that the exceptionally high thermal conductivity of the heat pipe is not solely exploited for transporting large quantities of heat, but also for obtaining surfaces of

in such a way that an optimum may be found by varying certain parameters. Others, such as the absolute operating range of a particular working fluid, must be regarded as fixed. Some of these limitations are shown in *fig. 2*. Clearly, a fluid is useless at temperatures below its melting point, since it can then no longer be recycled through the wick. At the other extreme, the latent heat of vaporization vanishes as the critical temperature is approached, and so a large part of the heat transport capability is lost. In general a fluid is considered 'interesting' only in the temperature range that corresponds to vapour pressures between say 0.1 and 10 atmospheres; for example, 650-1250 °C for sodium, 45-180 °C for water and −60-27 °C for ammonia.

The performance of heat pipes is subject to certain other limitations. In the first place, during 'start-up' difficulties may arise from the fact that the vapour stream has a maximum velocity related to the velocity
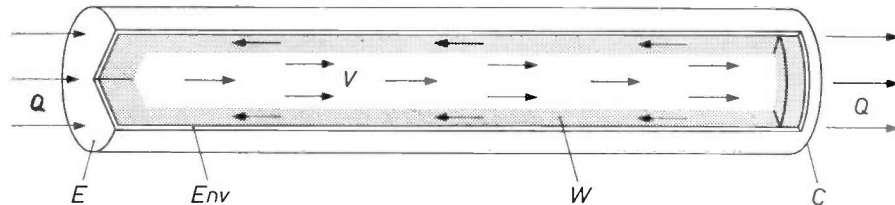
Fig. 1. Simple cylindrical heat pipe. The wall *Env* of the completely sealed pipe is clad on the inside with a porous layer *W*, the 'wick'. The heat entering the left-hand end of the pipe, the 'evaporator' *E*, causes fluid in the wick to evaporate. Latent heat is thus taken up by the vapour, and the vapour pressure at that end rises slightly. The vapour flows from the evaporator through the vapour duct *V* to the condenser *C*, where the latent heat is released upon condensation. The condensate is then recycled through the wick to the evaporator. In addition to the locations of the evaporator and the condenser — which need not be restricted to the ends — the pipe is provided with external heat insulation. *Q* input and output heat flux.

equal temperature. We shall return to this point in the section dealing with the theory of the thermal resistance of a heat pipe. Some applications based on this will be dealt with in part II of this article.

In general, the temperature drop across a heat pipe is determined by the amount of heat it is required to transport, while the mean operating temperature, that is the average temperature of the vapour in the vapour duct, is determined primarily by the temperature of the heat source or heat sink, whichever of the two is specified. When neither are fixed, as is the case when both heat input and heat output take place by a radiation process, then the heat pipe will adjust its operating temperature in accordance with its heat rejection capability. This generally means that the greater the required heat transport, the higher the mean operating temperature.

In practice, however, the heat transport through a heat pipe cannot be increased indefinitely. Some of the limitations depend upon the actual design of the pipe,
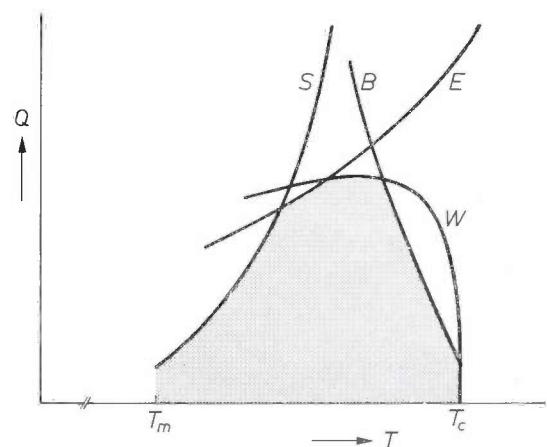
Fig. 2. The working range of a heat pipe. The axial heat flux *Q* is shown as a function of the temperature *T*. The heat pipe operates in the shaded area. The limits of *T* are the melting point T_m and the critical point T_c of the working medium. The upper limits of the heat transport in a heat pipe are determined in the first place by the velocity of sound in the vapour (curve *S*), and at higher operating temperatures by the maximum liquid-flow rate which the wick is capable of handling (curve *W*). Other limitations are due to the take-up of liquid from the wick by the vapour flow (curve *E*) and boiling of the fluid in the wick (curve *B*).

of sound in the vapour. Generally speaking, this 'sonic limit' as it is termed, is of importance only at low vapour pressures. The vapour density is then small as well, and the mass transport is thus severely limited. At higher temperatures and pressures, two other limits may be encountered, depending on the actual design of the heat pipe. The first of these is termed the *entrainment limit* and refers to the ability of high-velocity vapour streams to take up liquid from the wick and return it, in the form of high velocity droplets, to the condenser section. Clearly, wicks consisting of wide, open channels are more susceptible to this limitation than those with smaller pores. The second is the *wicking limit*, and refers to the ability of the wick to return liquid to the evaporator at the same rate as the liquid is being evaporated. A failure to sustain this supply results in burn-out of the evaporator, frequently with catastrophic consequences.

Another limitation, applicable at even higher operating temperatures, is the occurrence of boiling in the wick. As a result of this, vapour may block the supply of returning liquid and thus cause burn-out.

The relative positions of the limits are not fixed. For example, when a heat pipe is required to operate against a gravitational pressure head, that is with the evaporator *above* the condenser, almost all the operating range may be limited by the wicking capability. This applies particularly to working fluids of moderate temperature, which tend to have rather low surface tensions, a fact which has put an end to quite a number of proposals for heat pipe applications.

Heat-transport capability

The large heat-transport capability of heat pipes that work with liquid metal is almost legendary. It should be remembered, however, that at lower temperatures the maximum recorded heat fluxes are much more moderate. One reason for this is that working fluids of lower boiling point tend to have lower latent heats, in accordance with Trouton's rule. Some examples are given in *fig. 3*.

The high heat-transport capability of a heat pipe is due particularly to the transport of vapour along the vapour duct; the thermal resistance of the vapour duct is very small, as will be seen in the next section. The heat flux is in fact limited by other parts of the cycle, such as heat transfer into and out of the pipe, evaporation and condensation processes, etc. Moreover, the magnitude of the heat transfer capability should be judged in relation to the associated temperature gradient.

It is interesting to compare the performance of a heat pipe with that of other energy-transport systems. For example, it has been reported that a lithium
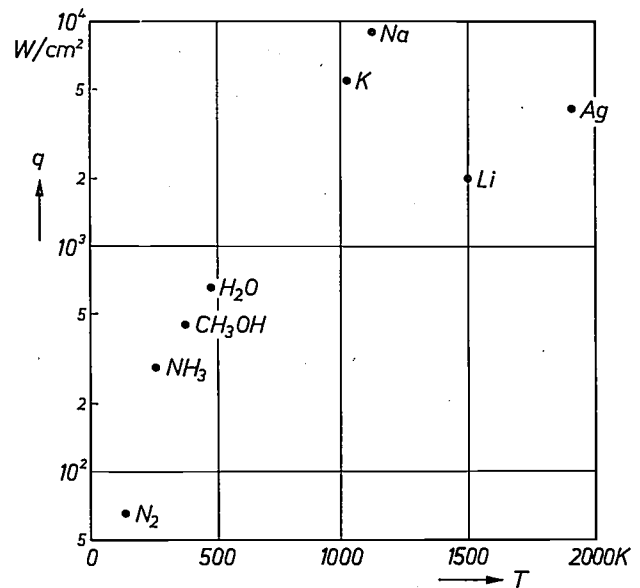
Fig. 3. Diagram of the measured maximum heat flux q for various working fluids, and the temperature at which these are ordinarily used. A higher operating temperature corresponds in general to a higher heat-transport capability.

heat pipe operating at 1500 °C can transport an energy of 15 kilowatts per square centimetre of cross-section with an axial temperature gradient of about 0.1 °C cm^{-1}. In comparison, a copper rod of the same dimensions transferring heat by conduction at a heat flux of 15 kilowatts cm^{-2} would require a temperature gradient of approximately 4000 °C cm^{-1}. Although the same comparison has frequently been made, it is perhaps a little unfair to the copper rod since it does not take into account the variation of the latter's properties with temperature. In the cryogenic region for example, at temperatures around 10 K, a pure copper rod can in principle handle 15 kilowatts cm^{-2} with a temperature gradient of only 75 °C cm^{-1}, and is probably as good a thermal conductor as any heat pipe operating in this temperature range.

It is also interesting to make a comparison between a heat pipe and a hollow pipe of the same dimensions, along which heat is exchanged between black-body thermal radiators. A simple calculation shows that if one end of the pipe could be maintained at absolute zero the other would have to be brought to approximately 6900 °C before the radiation flux approached 15 kilowatts cm^{-2}. Moreover, to obtain this energy flux for a temperature difference between the two ends of the same order as that encountered in heat-pipe systems, i.e. 5 °C, the mean operating temperature of such a thermal radiation pipe would have to be about 50 000 °C.

Although, as the above examples show, radiation and conduction processes barely stand comparison

with the heat-pipe process, there are, of course, means of energy transport that can. For example, the liquid sodium-potassium loops employed in the heat exchangers of nuclear reactors are quite effective. These operate by the absorption of sensible heat at the heat source, thus raising the temperature of the liquid alloy, which is then pumped to the heat sink where the sensible heat is dissipated and the alloy thus cooled. In this way, 15 kilowatts can be transferred along a loop with a cross-sectional area of one square centimetre, given that the difference between the temperature of source and sink is 200 °C and that the alloy is circulated with a velocity of about one metre per second.

A more bizarre example is the firing of bullets with a mass of 10 grams along the pipe. A repetition rate of 3 bullets per second brings one into the region of an energy transfer capability of about 15 kilowatts cm^{-2}. Similarly there need be no problems in the transfer of 15 kilowatts of electrical energy along a copper rod of 1 cm^2 cross-section. It is thus only in the transfer of

The high heat-transfer capability of the heat pipe therefore relies on two factors. The first is that incoming thermal energy may be converted to a potential energy and vice versa. The second is that the pressure drop associated with the flow of this potential energy in the form of vapour is small, as must also be the corresponding temperature difference between the evaporator and the condenser.

Thermal resistances in a heat pipe

To explain how the various useful characteristics of a heat pipe arise, it is convenient to consider a heat-pipe system as a series of thermal resistances, each originating from a distinct physical process in the transfer of heat from source to sink. At least nine such resistances may be identified, and these add up to give the total thermal resistance R_{tot} of the system:

$$R_{tot} = \Delta T/Q = \sum_{n=1}^{9} R_n = Q^{-1} \sum_{n=1}^{9} \Delta T_n.$$
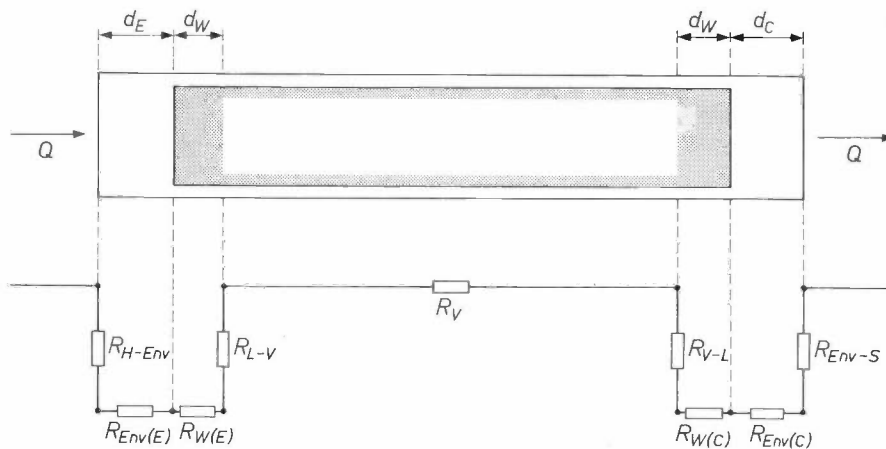
Fig. 4. The thermal resistance of a heat pipe can be regarded as the sum of a series of nine thermal resistances, each corresponding to a part of the pipe or to an interface. R_{H-Env} is the heater-envelope interface resistance. $R_{Env(E)}$ is the resistance of the envelope at the position of the evaporator. $R_{W(E)}$ is the resistance of the wick in the evaporator. R_{L-V} is the liquid/vapour interface resistance. R_V the resistance to the vapour flow; R_{V-L}, $R_{W(C)}$, $R_{Env(C)}$ and R_{Env-S} are resistances corresponding to the first four resistances (in the reverse order) but now relating to the condenser. At the location of the evaporator (left) and the condenser (right) the representation of the wall and the wick is exaggerated. The wall thicknesses are indicated by d_E, d_C and d_W respectively.

heat that the heat pipe shows a clear advantage, and the reason for this should be quite obvious. Like the bullet and the electrical conductor, the heat pipe does not primarily transport thermal energy. Certainly, thermal energy enters the heat-pipe system and thermal energy leaves it, but in between, in the transport stage, the energy is in the form of latent heat, which is the potential energy gained by a molecule on its removal from the liquid phase.

The nature of these several resistances is illustrated in *fig. 4.*

Apart from the heat flux itself, heat-pipe systems are substantially symmetrical in their characteristics: it is therefore only necessary to derive expressions for five of these resistances to determine which contribution to the total thermal resistance is the largest. In the following sections we shall take a look at the relevant theory. At the end of the section we shall then compare

the contributions of each of these resistances, and we shall see that three of the nine are so small as to be negligible.

Heater/envelope interface

The majority of heat pipes derive their heat input from a hot gas, a hot solid or a radiation source. We shall now consider each of these sources in turn.

Where the source is a hot gas, as is the case with flame-heated heat pipes, the heat flux per unit area of the heat-pipe envelope is determined by a relation of the form:

$$q = h\Delta T.$$

Then the thermal resistance of the heater/envelope interface is given by:

$$R_{\text{H-Env}} = (hA_{\text{E}})^{-1}.$$

In these expressions ΔT is the temperature difference between gas and envelope, h is the heat-exchange coefficient, and A_{E} is the surface area of the evaporator.

As an example of heating where a hot solid is the source, we consider an electric heating wire which is separated from the heat-pipe envelope by some form of electrical insulation. In this situation the heat-conduction equation applies:

$$q = \lambda\Delta Td^{-1}.$$

The associated thermal resistance is

$$R_{\text{H-Env}} = d(\lambda A_{\text{E}})^{-1}.$$

In these expressions ΔT is the temperature difference between heat-pipe envelope and heating wire, d is the thickness of the insulation and λ its thermal conductivity.

Where the heat source is a radiation source the heat is generated directly on the surface of the heat-pipe. There is then no heater/envelope interface so that $R_{\text{H-Env}} = 0$, as it is also, for example, in the case of electron bombardment heating or friction heating.

Heat transport through the envelope

The transport of heat through the heat-pipe envelope is for the most part primarily a process of conduction. Even when thermal radiation does make a contribution, the heat flow may still be expressed in terms of the general heat-conduction equation:

$$q = \lambda_{\text{E}}\Delta Td_{\text{E}}^{-1} \quad \text{and} \quad R_{\text{Env(E)}} = d_{\text{E}}(\lambda_{\text{E}}A_{\text{E}})^{-1}.$$

These expressions indicate that thin-walled, high-conductivity materials are to be preferred for heat-pipe envelopes. Unfortunately the need to contain pressures usually puts a lower limit on the thickness of the en-

velope; and the possibility of corrosion imposes restrictions on the type of material that can be employed.

Transport through the wick

The transport of heat through a liquid-saturated wick may occur in several ways, depending upon the nature of the wick and the working medium. For liquid metals the largest contribution is likely to arise from conduction through the liquid itself. On the other hand, with an organic fluid and a metallic wick, conduction through the wick may assume the dominant role. If the wick is fabricated from an insulating material such as fibre-glass, then convection or even boiling in the liquid become possible transport mechanisms.

In many situations an analysis of the problem is not possible, and therefore it is useful to have an estimate of the upper limit to the thermal resistance of the wick. This may be obtained by assuming that all the heat is transported solely by conduction through the liquid, in which case the resistance is given by:

$$R_{\text{W(E)}} = d_{\text{W}}(\lambda_{\text{W}}A_{\text{E}})^{-1}.$$

Liquid/vapour interface

The thermal resistance at the liquid/vapour interface is not so easily treated as the foregoing resistances. We begin by considering a heat pipe in static equilibrium, that is to say a heat pipe not transporting heat. This is roughly the situation of a heat pipe used for obtaining an isothermal surface. The temperatures of the liquid and the vapour will then be the same everywhere and equal to the temperature of the immediate surroundings. The pressure of the vapour is therefore exactly equal to the vapour pressure of the liquid at that temperature. Now the vapour pressure is in general an exponential function of temperature, as may be deduced from the Clausius-Clapeyron equation:

$$dP/dT = \frac{L}{T(V_{\text{v}} - V_{\text{L}})}.$$

In this expression L is the latent heat, V_{V} is the volume of the vapour and V_{L} is the volume of the liquid.

Assuming that V_{L} is negligible in comparison with V_{V}, we may write: $dP/dT = L(V_{\text{v}}T)^{-1}$. Assuming further that the vapour may be treated as a perfect gas, we find $dP/dT = LP(RT^2)^{-1}$, R being the gas constant. A solution of this equation is:

$$P \propto \exp\left(\frac{-L}{RT}\right).$$

From this it is evident that the first derivative dP/dT increases rapidly with increasing temperature, so that at higher temperatures relatively small changes of temperature give rise to quite drastic changes in vapour

pressure. This feature is extremely important in applications requiring good temperature stability.

For a heat pipe in the dynamic state, but transporting *little* heat, the Clausius-Clapeyron equation provides a means of relating the small pressure drops to equivalent temperature drops, i.e.

$$\Delta P = \Delta T\, LP(RT^2)^{-1}.$$

For a heat pipe transporting a *large* amount of heat, the situation is slightly different. At the interface liquid is vaporizing and the heat input is thus converted into potential energy (latent heat). To obtain finite vaporization rates, some degree of superheating of the liquid is required. The heat-transfer rate may be expressed by a relation of the form

$$q = \alpha \Delta T,$$

where ΔT is the temperature difference between liquid and vapour, and α is the heat-transfer coefficient, which is usually experimentally determined. For water evaporating from a steel-gauze wick, α is typically of the order of 0.5 watts cm^{-2} °C^{-1}, with burn-out occurring at about 10 watts cm^{-2}. Observations [3] made on room-temperature heat pipes indicate that whether vaporization occurs with or without boiling, the degree of superheating required to effect a given heat-flux transfer from a wick is, at the lower levels of heat flux, less than that encountered in normal pool boiling, but increases above the pool-boiling value at higher heat fluxes.

With a liquid like water, the superheating required for boiling is only a few degrees; as far as liquid metals are concerned, such as sodium and potassium, superheating to between 100 and 200 degrees might be necessary.

A relation between heat flux and superheat such as that given above provides little quantitative information on the effectiveness of vaporization as an aid to heat transfer. Another approach may be found by noting that when a heat pipe is in the dynamic state, i.e. transporting heat, the vapour pressure of the liquid at the evaporator must be slightly higher than the pressure of the vapour there; whereas at the condenser, finite condensation rates are obtained only when the pressure of the vapour is slightly higher than the vapour pressure of the liquid. This is one of the consequences that may be deduced from the kinetic theory of gases, which in fact offers a good description of both evaporation and condensation phenomena.

Consider a surface element of unit area parallel and very close to the surface of the liquid. The vapour below this surface may be taken as having the equilibrium vapour pressure P of the liquid (temperature T) and is in dynamic equilibrium with the vapour above,

at pressure P_V. The number of molecules travelling from the liquid to the vapour side of the element, per second and per unit area, is from kinetic theory

$$N_{L-V} = (3/2\pi)^{\frac{1}{2}}(P/mc),$$

while the number travelling in the opposite direction is

$$N_{V-L} = (3/2\pi)^{\frac{1}{2}}(P_V/mc).$$

In these expressions L-V means liquid to vapour, and V-L vapour to liquid; m is the mass of a molecule of the vapour, and c is the root-mean-square velocity, a quantity defined in terms of the average kinetic energy per molecule, i.e.

$$c = (3kT/m)^{\frac{1}{2}} = (3RT/M)^{\frac{1}{2}},$$

where k is Boltzmann's constant and M the gram-molecular weight of the vapour. The net number of molecules crossing unit area per second is therefore:

$$N = (3/2\pi)^{\frac{1}{2}}(P - P_V)/(mc).$$

The rate of heat transfer is, to a good approximation,

$$q = NmL = (P - P_V)L(M/2\pi RT)^{\frac{1}{2}}.$$

The rate at which heat is transferred across a liquid/vapour interface is therefore determined not only by the latent heat of the liquid, but also by the difference between the pressure of the vapour and the vapour pressure of the liquid. The amount of heat transferred for only a small pressure difference can be quite surprising: some values of $q/(P - P_V)$ estimated for several liquids in the vicinity of their boiling points are given in *Table I*.

Using the previously derived result relating temperature and vapour pressure differences (p. 108), the rate of heat flux across the liquid/vapour interface may be written as

$$q = \Delta T\, L^2 P(M/2LRT)^{\frac{1}{2}}(RT^2)^{-1}$$

and the thermal resistance of the liquid/vapour interface is

$$R_{L-V} = (RT^2/L)(2\pi RT/M)^{\frac{1}{2}}(LPA_E)^{-1}.$$

Table I. The boiling point T_b of some working fluids for heat pipes, and the ratio of the heat flux q through the surface of an evaporating liquid and the difference which is required between the vapour pressure P_V and the saturation pressure P.

Liquid	T_b (K)	$q/(P - P_V)$ (kW/cm^2 atm)
Lithium	1613	55
Zinc	1180	18
Sodium	1156	39
Water	373	21.5
Ethanol	351	13.5
Ammonia	238	15.2

[3] P. J. Marto and W. L. Mosteller, ASME-AIChE Heat Transfer Conf., Minneapolis 1969, paper No. 69-HT-24.

The assumptions made in this derivation are not tenable for large pressure differences, but it is probable that in any real situation these differences will be very small, at least in comparison to the absolute pressure in the pipe.

Vapour flow

Clearly, vapour flows from the evaporator to the condenser only in the presence of a pressure gradient along the vapour duct. The magnitude of the pressure drop arising from frictional losses in the vapour flow has been calculated by T. P. Cotter [4] for various situations. The result obtained in the limit of low evaporation and condensation rates is

$$\Delta P_V = 4\eta_V lq/(L\varrho_V r_V^4).$$

Here η_V is the viscosity of the vapour, l the length of the heat pipe, ϱ_V the density of the vapour, and r_V the radius of the vapour duct. As before, a temperature drop may be related to this pressure drop via the Clausius-Clapeyron equation. The corresponding thermal resistance arising from frictional losses in the vapour flow is thus

$$R_V = (RT^2/L)(4\eta_V l)(P_V L\varrho_V r_V^2 A_V)^{-1},$$

where A_V is the cross-sectional area of the vapour duct.

The other four thermal resistances have essentially the same form as the first four and will not be discussed further, although they are included in *Table II*.

Table II. The expressions for the nine thermal resistances that can be distinguished in a heat pipe (see text), and approximate values per cm² of cross-section for water as the working fluid.

Resistance	Expression	Approx. value (°C/W)
R_{H-Env}	$(hA_E)^{-1}$ or $d(\lambda A_E)^{-1}$	10^3 or 10
$R_{Env(E)}$	$d_E(\lambda_E A_E)^{-1}$	10^{-1}
$R_{W(E)}$	$d_W(\lambda A_E)^{-1}$	10
R_{L-V}	$(RT^2/L)(2\pi RT/M)^{\frac{1}{2}}(LPA_E)^{-1}$	10^{-5}
R_V	$(RT^2/L)(4\eta_V l)(P_V L\varrho_V r_V^2 A_V)^{-1}$	10^{-8}
R_{V-L}	$(RT^2/L)(2\pi RT/M)^{\frac{1}{2}}(LPA_c)^{-1}$	10^{-5}
$R_{W(C)}$	$d_W(\lambda_W A_c)^{-1}$	10
$R_{Env(C)}$	$d_C(\lambda_C A_c)^{-1}$	10^{-1}
R_{Env-S}	$(hA_c)^{-1}$ or $d(\lambda A_c)^{-1}$	10^3 or 10

Table II shows thermal-resistance values calculated for a hypothetical water-filled heat pipe. It is seen that the internal resistances associated with the mass-flow phenomena, that is R_{L-V}, R_V and R_{V-L} are negligible compared with those of the envelope, wick and heat input and output systems.

The total thermal resistance of a heat-pipe system may thus be estimated from consideration of the latter six resistances alone; these will in general be sufficient for determining the heat transported between a source and sink at fixed temperatures, or for determining the temperature drop expected for a given heat-transfer rate.

Design and performance

In the design of a heat pipe the working fluid, the envelope and wick materials, the dimensions and wick structure must of course be chosen to suit the operating conditions. The final choice is invariably the result of some sort of compromise. Some of the relevant parameters and criteria are discussed below.

Pressure drops in a heat pipe

One approach to recognizing the various parameters that may come into conflict in the design of a heat pipe is to consider the pressure differences that must occur in a heat pipe. There are four such pressure differences — the capillary pumping pressure ΔP_{cap}, the hydrostatic pressure ΔP_{grav}, the pressure drop in the liquid ΔP_L and the pressure drop in the vapour ΔP_V — whose sum must be zero:

$$\Delta P_{cap} + \Delta P_{grav} + \Delta P_L + \Delta P_V = 0.$$

Each of these terms has been evaluated for the case of a simple cylindrical heat pipe, absorbing heat uniformly along the length of its evaporator, and liberating this heat uniformly along the length of its condenser [4].

The capillary pumping pressure ΔP_{cap} is due to the curvature of the liquid/vapour interface, one of which exists at both ends of the heat pipe. Let the radii of curvature be R_E and R_C and the surface tension of the liquid be γ, then the net pressure difference is $\Delta P_{cap} = 2\gamma(R_E^{-1} - R_C^{-1})$. However, because of the dynamic nature of the condensation process R_C^{-1} may, generally, be treated as negligible with respect to R_E^{-1}. Moreover, a simple geometrical construction reveals that $R = r_c/\cos\theta$, where r_c is the radius of a capillary, and θ is the contact angle between liquid and capillary. A sufficiently good approximation for the capillary-pressure term is therefore:

$$\Delta P_{cap} \leqslant (2\gamma/r_c)\cos\theta.$$

The second term, ΔP_{grav}, the hydrostatic pressure drop in the liquid, is given by $\Delta P_{grav} = \varrho_L gl\sin\phi$, where ϱ_L is the density of the liquid, l the length of the heat pipe and ϕ its inclination to the horizontal. This term is positive or negative, depending on whether it helps or hinders the recycling of the liquid.

The hydrodynamic pressure drop in the wick, ΔP_L, may be derived from the Poiseuille formula, corrected for tortuosity and wall roughness. The result may be written in the form $\Delta P_L = -\dot{M}lb\eta_L/(2A_W\varrho_L\varepsilon r_c^2)$, where $\dot{M}$ is the mass flow rate, b a factor representing

tortuosity and roughness, η_L the viscosity of the liquid, A_W the cross-sectional area of the wick and e its porosity.

A similar result is obtained by application of the general pressure-drop equation:

$$\Delta P_L = - R^* C_W l v_L \eta_L / 8 r_c^2,$$

where R^* is the axial Reynolds number of the liquid, v is the velocity of the fluid and C_W is the coefficient of friction. Comparison with the previous result requires that $b = R^* C_W / 8$. The product $R^* C_W$ is 64 for straight, smooth capillaries, 96 for channels, and 100-200 for gauzes, depending on their filling factors.

The pressure drop ΔP_V in the vapour is related to frictional losses in the vapour flow, and also to the fact that the axial component of the velocity of molecules that have just transferred from the liquid to the vapour phase is zero on average, whereas at the output of the evaporator it has quite a high value. This pressure difference is thus primarily an inertial effect. Analogous considerations apply to the condenser. If the rates of evaporation and condensation are high, the contributions from the inertial effects may predominate over the frictional losses. In the limit case of low evaporation and condensation rates, frictional forces dominate, and the pressure drop ΔP_V in the vapour is approximately equal to $-4\eta_V \dot{M} l (\pi \varrho_V r_V^4)^{-1}$, where η_V is the viscosity of the vapour, ϱ_V its density and r_V the radius of the vapour duct. The total pressure drop in the vapour for high heat fluxes is $\Delta P_V = -\dot{M}^2 (1 - 4/\pi^2)/8\varrho_V r_V^4$. The first term is the pressure drop in the evaporator part and the second is that in the condenser part.

For high evaporation rates, that is high radial flow velocities, the velocity profile across the vapour duct is not parabolic but a cosine function. At high condensation rates the velocity profile in the condenser is nearly constant across the vapour duct with the transition to zero velocity occurring in a thin layer near the wall. Surprisingly, the pressure in the condenser actually increases in the direction of flow [5].

Using the above expressions for ΔP_{cap}, ΔP_{grav}, ΔP_L and ΔP_V, and substituting Q/L ($Q =$ heat flux, $L =$ latent heat) for $\dot{M}$, we obtain from the requirement that the sum of the pressure drops must be zero, the following equations. For the case of a low heat input/output:

$$\frac{4\eta_V Q l}{\pi \varrho_V r_V^4 L} + \varrho_L g l \sin \phi + \frac{b\eta_L Q l}{2 A_W \varrho_L r_c^2 L e} \leqslant \frac{2\gamma \cos \theta}{r_c}.$$

and for the case of a high heat input or output:

$$\frac{(1 - 4/\pi^2) Q^2}{8\varrho_V r_V^4 L^2} + \varrho_L g l \sin \phi + \frac{b\eta_L Q l}{2 A_W \varrho_L r_c^2 L e} \leqslant \frac{2\gamma \cos \theta}{r_c}.$$

The above equations can be quite a useful aid to heat-pipe design, even though they are strictly applicable only in special cases. Nevertheless, they give some insight into the magnitudes that various parameters, such as the cross-sectional area of the vapour duct or the capillary radius of the wick, must have if the heat pipe is to operate well. Certain parameters, such as the porosity e and the tortuosity of the wick, usually have to be derived from experiment.

Fundamental limitations are particularly likely to be encountered in the design of heat pipes that are required to transport a great deal of heat. For example, flow losses in the wick increase as the diameter of the pores decreases, which may offset the gain in capillary pressure. This effect restricts the length; it has not yet proved possible to give any heat pipe, required to transport a large amount of heat, a greater length than a few metres, without adding an auxiliary system to maintain the liquid flow.

Choice of working fluid

We consider a heat pipe in which the pressure drop in the vapour is small enough for it to be neglected in comparison with the pressure drops occurring in the liquid phase. We further assume that the capillaries are perfectly wetted, i.e. $\theta = 0$. Under these conditions the sum of the pressure drops occurring in the heat pipe is approximately given by

$$\frac{b\eta_L Q l}{2 A_W \varrho_L r_c^2 L e} + \varrho_L g l \sin \phi - \frac{2\gamma}{r_c} = 0,$$

or

$$Q = \left[\frac{2\gamma}{r_c} - \varrho_L g l \sin \phi \right] \frac{2 A_W \varrho_L r_c^2 L e}{b\eta_L l} =$$

$$= \frac{4 A_W e r_c}{bl} \left[1 - \frac{\varrho_L}{\gamma} \frac{g l r_c}{2} \sin \phi \right] \frac{L \gamma \varrho_L}{\eta_L}.$$

As can be seen, the expression for Q is a product of three factors. The first two are primarily concerned with the geometry of the wick and the orientation of the pipe. The last term is the 'figure of merit' F of the working fluid. It includes the most important physical properties relevant to heat-pipe performance. Clearly, for high heat-transport capability we must seek fluids with a high figure of merit.

Fig. 5 gives some examples of the figures of merit of several fluids, estimated at their boiling points. The range of values is very large and shows a definite trend to higher values for higher-boiling-point liquids. For a given liquid, F is a temperature-dependent property which must vanish at the critical point. In general,

[4] T. P. Cotter, Theory of heat pipes, Los Alamos Sci. Lab. Report No. LA-3246-MS; 1965.
[5] W. E. Wageman and F. A. Guevara, Phys. Fluids 3, 878, 1960.
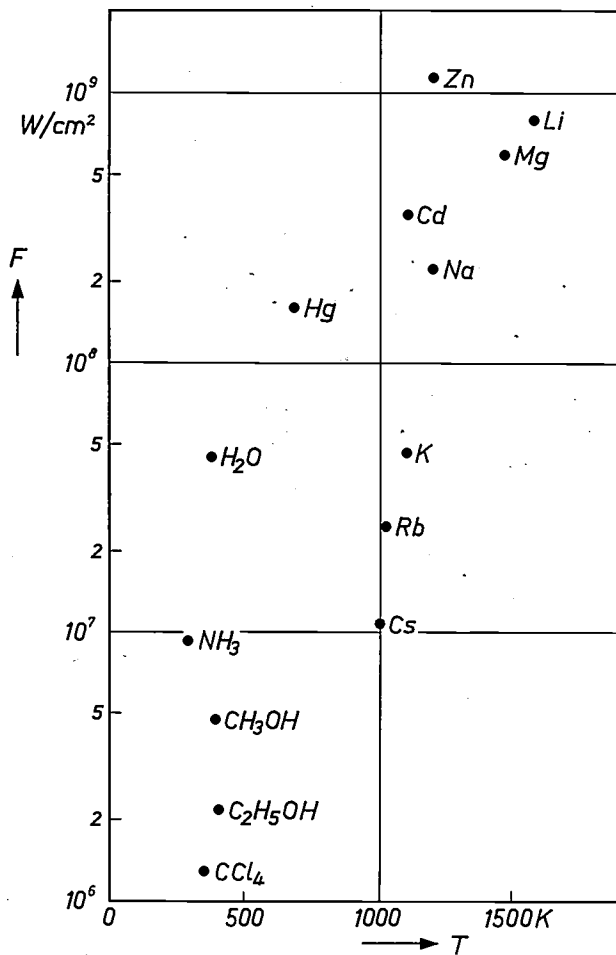
Fig. 5. The figure of merit F for various working media at their boiling points.

Table III. Properties of some working media. T_m melting point. T_b boiling point. ϱ_L density, γ surface tension and F figure of merit at boiling point.

Medium	T_m (K)	T_b (K)	ϱ_L (g/cm³)	γ/ϱ_L (cm³/s²)	F (W/cm²)
NH₃	195.3	239.7	0.685	38	976×10^4
CH₃OH	179	338	0.78	22	486×10^4
CCl₄	250	349.5	1.5	13	116×10^4
C₂H₅OH	156	351.5	0.75	23	215×10^4
H₂O	273	373	0.96	61	456×10^5
Naphthalene	353.5	491	0.89	22	112×10^4
Hg	234.1	629.5	12.74	30	162×10^6
Rb	311.9	961	1.2	38	250×10^5
Cs	301.5	958	1.47	22	124×10^5
Cd	593.9	1038	7.6	73	376×10^6
K	336.7	1047	0.656	91	450×10^5
Na	370.8	1156	0.756	153	227×10^6
Zn	692.6	1180	6.2	118	114×10^7
Mg	924	1380	1.5	347	590×10^6
Li	452	1613	0.404	530	825×10^6

values tend to increase from the triple point, reach a maximum around the normal boiling point of the liquid, and thereafter decreases quite rapidly as the critical point is approached. *Table III* lists a number of working fluids and their properties.

The figure of merit is not, of course, the only factor involved in the choice of a working medium for a heat pipe. In fact, there are instances when this parameter is not at all decisive. Four such instances are touched on below.

When the inclination of the pipe is steep, a high capillary rise is required in the wick. In this case, low-density fluids with high surface tensions (high γ/ϱ_L, see Table III) and small contact angles are preferable. The wick must also be chosen with as much care as the working medium.

In situations where great pressure differences across the heat-pipe envelope cannot be tolerated, working fluids should be chosen which, at the required operating temperature, have vapour pressures similar to the ambient pressure.

In some cases the fluid that has the most favourable figure of merit is likely to cause excessive corrosion of the heat pipe. For example, at 1400 K lithium has a much higher figure of merit than most other liquid metals. At this temperature, however, lithium attacks stainless steel quite severely. It is therefore more convenient in practice to employ sodium with an over-pressure than to make the heat-pipe envelope from some special lithium-resistant alloy, even though the vapour pressure of sodium at 1400 K is about 3 atmospheres, which brings special problems in designing the wall since the creep strength of stainless steel is low at this temperature.

The working fluid with the most superior figure of merit may also have to be rejected if the application requires from it some subsidiary property, such as a low electrical conductivity.

Finally, an important non-technical factor to be considered is the price of the working medium.

The wall and the wick

For the envelope material the requirements of high heat-transport capability and small temperature gradients lead naturally to materials with high thermal conductivities. For this reason copper is very suitable for use at temperatures below 300 °C, especially so on account of its availability in tubular form. At higher temperatures, where there is likely to be a problem from corrosion due both to impurities in the working fluid and the ambient atmosphere, we have made extensive use of two types of chrome-nickel steel (AISI 310 and 321). Where possible, seams are joined by argon-arc welding, although for temperatures up to 1000 °C nickel-based brazing has also been successfully employed. A sodium heat pipe fabricated with these techniques, which has operated for 30 000 hours, continuously is shown in *fig. 6*.

To make the use of heat pipes an economic prop-

osition up to temperatures above 1200 °C cheap wall materials will have to be developed. Refractory metals, such as tantalum, and certain alloys such as niobium-zirconium and tungsten-renium, have been employed by others [6], but construction with these materials is

wicks, i.e. wicks with a small number of relatively wide channels, are to be preferred.

The amount of working fluid required in a heat pipe is determined from knowledge of the wick volume and its porosity.
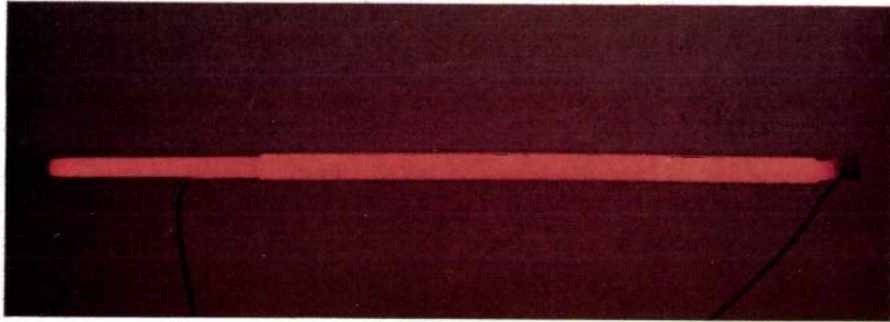
Fig. 6. A heat-pipe during a life test. This small sodium-filled heat pipe, made of stainless steel (30 cm long) has operated for more than 30 000 hours continuously — i.e. nearly $3\frac{1}{2}$ years — at a temperature of 750 °C.

not at all simple. It seems that there might also be some chance of success in producing ceramics able to withstand the alkali and alkaline-earth metals.

We have found it in general most convenient to fabricate the wicks from several layers of stainless-steel gauze spot-welded to the heat-pipe wall. This type of wick is only really suitable, however, for low-performance heat pipes, like those for isothermal ovens. If higher liquid-transport rates are required, arterial

Summary. In its simplest form a heat pipe is a closed duct clad on the inside with a porous layer, the 'wick', in which a liquid is contained. When one end of the pipe is heated, the liquid evaporates there. The vapour transports a very large part of the absorbed heat in the form of latent heat, which is given up at the other end of the pipe, where the vapour condenses. The condensate is recycled through the wick. A heat pipe has an exceptionally high thermal conductivity, and can therefore be used both for transporting very large quantities of heat and for obtaining isothermal surfaces. The total thermal resistance can be regarded as a series of nine resistances, whose characteristics are analysed. This analysis and the theory of the pressure drops in the various parts of the pipe lead to an expression that yields a figure of merit for the working medium. This is: $L\gamma\varrho_{L}/\eta_{L}$, where L is the latent heat of the working medium, γ the surface tension ϱ_{L} the density and η_{L} the viscosity. The wall and the wick are also required, of course, to meet specific requirements. A stainless-steel heat pipe operating at 750 °C with liquid sodium as the working medium has now operated continuously for about $3\frac{1}{2}$ years.

[6] See for example: C. A. Busse, Heat pipes for thermionic space power supplies, 3rd Int. Conf. of Space Technology, Rome 1971.

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, Netherlands	E
Mullard Research Laboratories, Redhill (Surrey), England	M
Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes (Val-de-Marne), France	L
Philips Forschungslaboratorium Aachen GmbH, Weißhausstraße, 51 Aachen, Germany	A
Philips Forschungslaboratorium Hamburg GmbH, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium.	B

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, Netherlands.

C. C. J. Addink & G. Meijer: Kinetic studies on the auxin effect and the influence of cycloheximide and blue light.
Plant growth substances 1970, editor D. J. Carr, pp. 68-75; publ. Springer, Berlin 1972. E

H. Bex: Theorie des magnetisch abstimmbaren Bandpaßfilters.
Nachrichtentechn. Z. 25, 390-394, 1972 (No. 9). A

P. Blood: Electrical properties of n-type epitaxial GaAs at high temperatures.
Phys. Rev. B 6, 2257-2261, 1972 (No. 6). M

D. Boccon-Gibod, J. L. Teszner (Ecole Normale Supérieure, Paris) & M. Mautref: Modification of Gunn instabilities in coplanar diodes by variation of the contact depth.
Electronics Letters 8, 119-120, 1972 (No. 5). L

J. van den Boomgaard & L. R. Wolff: Transformation of M_7C_3 into $M_{23}C_6$ in fibre-reinforced M-M_7C_3 eutectic composite materials and the impact on the strength properties.
Philips Res. Repts. 27, 509-512, 1972 (No. 6). E

A. H. Boonstra & R. M. A. Sidler: Successive adsorptions of hydrogen bromide and hydrogen sulfide on vapor-deposited lead monoxide layers.
J. Electrochem. Soc. 119, 1193-1197, 1972 (No. 9). E

M. J. G. Braken: Micro-plasmaproeven met een mechanische toorts en een mechanische klembeweging.
Lastechniek 38, 204-207, 1972 (No. 9). E

P. C. Brandon: β-bromo-β-nitrostyrene: a new energy transfer inhibitor in chloroplast photophosphorylation.
Proc. 2nd Int. Congress on Photosynthesis, Stresa 1971, pp. 1015-1022; 1972. E

J. J. van den Broek: Variations in atomic volume for alloy systems.
Physics Letters 40A, 219-220, 1972 (No. 3). E

J. W. Broer: Autoritis, een gezond ziektebeeld.
Ned. T. Natuurk. 38, 249-254, 1972 (No. 15). E

K. H. J. Buschow: Note on the stability of rare earth - cobalt compounds with $CaCu_5$ structure.
J. less-common Met. 29, 283-288, 1972 (No. 3). E

K. H. J. Buschow & H. H. van Mal: Phase relations and hydrogen absorption in the lanthanum-nickel system.
J. less-common Met. 29, 203-210, 1972 (No. 2). E

F. M. A. Carpay: Gerichtete Verbundwerkstoffe durch Feststoff-Umwandlungen.
Chemie-Ing.-Technik 44, 878-881, 1972 (No. 14). E

J. Cornet & D. Rossier: Glass-forming ability vs local order and covalent character in the As-Te system.
Appl. Phys. Letters 20, 304-306, 1972 (No. 8). L

R. Davies, W. S. C. Gurney (University of Strathclyde, Glasgow) & A. Mircea: Frequency/temperature characteristics of Gunn devices.
Electronics Letters 8, 349-351, 1972 (No. 14). M, L

M. Davio: Ring-sum expansions of Boolean functions.
Proc. Symp. on Computers and Automata, New York 1971, pp. 411-418. B

K. Eckrich, E. Dormann, A. Oppelt (all with II. Phys. Inst. Techn. Hochschule Darmstadt) & K. H. J. Buschow: NMR investigations on Y- and La-diluted GdZn.
Physics Letters 41A, 23-24, 1972 (No. 1). E

G. Eschard, C. Loty, A. Pelissier & J. F. Bonnal: Basic studies of a cesium contact ion thruster.
AIAA 9th Electric Propulsion Conf., Washington 1972, AIAA paper No. 72-440, pp. 1-8. L

A. C. Everitt, E. D. Fletcher & D. H. Paxman: The performance of a Ho^{3+} in yttrium aluminium garnet laser system.
Proc. MOGA Conf., Amsterdam 1970, pp. 21.1-21.5. M

E. Fabre: Copper contamination during the vapour epitaxial growth of GaAs.
Phys. Stat. sol. (a) **9**, 259-262, 1972 (No. 1).   *L*

E. Fabre & C. Schiller: Correlation between the presence of dislocations and vacancies in *n*-type gallium arsenide.
Solid State Comm. **10**, 81-84, 1972 (No. 1).   *L*

R. C. French: Speech scrambling.
Electronics & Power **18**, 263-264, 1972 (July).   *M*

P. Gerthsen, K. H. Härdtl & A. Csillag: Mobility determinations from weight measurements in solid solutions of $(Ba,Sr)TiO_3$.
Phys. Stat. sol. (a) **13**, 127-133, 1972 (No. 1).   *A*

A. E. Giles (Hurstpierpoint College, Hassocks, Sussex) **& D. Washington:** An experiment in school-industry cooperation.
Physics Bull. **23**, 401-402, 1972 (July).   *M*

A. Godard: Modèle mathématique de la mécanique pulmonaire.
J. Inform. méd. IRIA, Rocquencourt, **2**, 119-129, 1972.   *B*

G. G. P. van Gorkom & A. T. Vink: ESR of a Mn-S nearest neighbour complex in GaP.
Solid State Comm. **11**, 767-770, 1972 (No. 6).   *E*

R. G. Gossink & J. M. Stevels: Viscosity of molten alkali molybdates, tungstates, and sulfates.
Inorg. Chem. **11**, 2180-2182, 1972 (No. 9).   *E*

P. Guétin & G. Schréder: Effects of an ion bombardment on the characteristics of a metal/*n*-GaAs tunnel contact.
J. appl. Phys. **43**, 549-561, 1972 (No. 2).   *L*

P. Guétin & G. Schréder: Tunneling in *n*-type-GaAs-Pb contacts under pressure.
Phys. Rev. B **5**, 3979-3988, 1972 (No. 10).   *L*

H. B. Haanstra: Elektronenmikroskopen.
Klei en Keramiek **22**, 185-189, 1972.   *E*

J. Hasker: Beam-current characteristic and cathode loading of electron guns with rotational symmetry: some important properties and method of calculation.
Philips Res. Repts. **27**, 513-538, 1972 (No. 6).   *E*

H. Haug & K. Weiss: A quantum mechanical determination of the superfluid density.
Phys. kondens. Mat. **14**, 324-335, 1972 (No. 4).   *E*

B. Hill: Breitbandige Nachrichtenvermittlung mit Laserstrahl-Koppelfeldern.
Nachrichtentechn. Z. **25**, 385-389, 1972 (No. 9).   *H*

K. Holford: Cat's eye radar.
SERT J. **6**, 109-111, 1972 (No. 5).   *M*

R. Hulin (Technische Universität Braunschweig) **& J. J. Goedbloed:** Influence of carrier diffusion on the intrinsic response time of semiconductor avalanches.
Appl. Phys. Letters **21**, 69-71, 1972 (No. 2).   *E*

Y. Kamp: Some open questions in the theory of non-commensurate distributed networks.
Network theory, editor R. Boite, publ. Gordon & Breach, London 1972, pp. 181-190.   *B*

E. T. Keve, K. L. Byë, P. W. Whipps & A. D. Annis: The effects of additives on the pyroelectric properties of T.G.S.
J. Physique **33**, C2/229-231, 1972 (Colloque C2).   *M*

C. Knobler, C. Romers (both with University of Leiden), **P. B. Braun & J. Hornstra:** The conformation of non-aromatic ring compounds, LXXV. The crystal and molecular structure of the 3,20-bis(ethylenedioxy) analogue of vitamin D.
Acta cryst. B **28**, 2097-2103, 1972 (No. 7).   *E*

A. J. R. de Kock & P. G. T. Boonen: A new method for revealing striations in high-resistive floating-zone silicon crystals.
J. Electrochem. Soc. **119**, 1241-1243, 1972 (No. 9).   *E*

J.-P. Krumme, P. Hansen & J. Haberkamp: Local magnetocrystalline and induced anisotropies in liquid-phase epitaxial garnet films.
Phys. Stat. sol. (a) **12**, 483-490, 1972 (No. 2).   *H*

H. K. Kuiken: On the flow between two independently rotating disks of variable distance with blowing.
Philips Res. Repts. **27**, 539-582, 1972 (No. 6).   *E*

F. K. Lotgering & G. H. A. M. van der Steen: Crystal structure and magnetic properties of $Cu_4Mn_2Te_4$.
J. Phys. Chem. Solids **33**, 2071-2078, 1972 (No. 11).   *E*

M. H. van Maaren: Superconductivity in tin - group V^a trichalcogenides.
Physics Letters **40A**, 353-354, 1972 (No. 5).   *E*

F. J. MacWilliams, N. J. A. Sloane (both with Bell Telephone Labs, Murray Hill, N.J., USA) **& J.-M. Goethals:** The MacWilliams identities for nonlinear codes.
Bell Syst. tech. J. **51**, 803-819, 1972 (No. 4).   *B*

R. Memming & F. Möllers: Two-step redox processes with quinones at semiconductor electrodes.
Ber. Bunsen-Ges. phys. Chemie **76**, 609-616, 1972 (No. 7).   *H*

H. Meyer, D. Riekmann, K. P. Schmidt, U. J. Schmidt, M. Rahlff (Elektro Spezial, Hamburg), **E. Schröder** (Elektro Spezial, Hamburg) **& W. Thust** (Elektro Spezial, Hamburg): Design and performance of a 20-stage digital light beam deflector.
Appl. Optics **11**, 1732-1736, 1972 (No. 8).   *H*

A. E. Mircea: Computer optimized design of pulsed Gunn oscillators.
IEEE Trans. **ED-19**, 21-26, 1972 (No. 1).   *L*

B. J. Mulder: Optical properties of crystals of cuprous sulphides (chalcosite, djurleite, $Cu_{1.9}S$, and digenite).
Phys. Stat. sol. (a) **13**, 79-88, 1972 (No. 1).   *E*

A. G. van Nie: Representation of linear passive noisy 1-ports by two correlated noise sources.
Proc. IEEE **60**, 751-753, 1972 (No. 6).   *E*

K. J. van Oostrum: On-line analysis of diffraction patterns.
Proc. 5th European Congress on Electron Microscopy, Manchester 1972, pp. 618-619.   E

O. J. Poppema, B. J. Verhaar, H. L. Hagedoorn (all with Eindhoven University of Technology) & E. Bruninx: Het cyclotronlaboratorium van de T.H.E.
Ned. T. Natuurk. 37, 55-63, 1971 (No. 4).   E

H. F. Premsela: Some aspects of the modulation transfer of the EM-TV chain.
Proc. 5th European Congress on Electron Microscopy, Manchester 1972, pp. 168-169.   E

W. Puschert: Automatic mode locking of cw lasers.
Optics Comm. 5, 380-381, 1972 (No. 5).   A

W. J. J. Rey: A noise-variance estimator in electro-encephalography.
Computers & biomed. Res. 5, 108-113, 1972 (No. 2).   B

W. Rey, R. Bernard (Hôpital Saint Pierre, Bruxelles), H. Vainsel (Hôp. St. Pierre, Br.) & A. Swietochowski (Hôp. St. Pierre, Br.): Surveillance électrocardiographique et détection par ordinateur des arythmies.
J. Inform. méd. IRIA, Rocquencourt, 2, 325-330, 1972.   B

F. Rondelez, H. Arnould & C. J. Gerritsma: Electro-hydrodynamic effects in cholesteric liquid crystals under ac electric fields.
Phys. Rev. Letters 28, 735-737, 1972 (No. 12).   L, E

P. Saraga & P. R. Wavish: Edge tracing in binary arrays.
Proc. Conf. on Machine Perception of Patterns and Pictures, Teddington 1972, pp. 294-302.   M

B. Schiek & K. Schünemann (Valvo, Hamburg): Comments on 'Noise in IMPATT diode amplifiers and oscillators'.
IEEE Trans. MTT-20, 631, 1972 (No. 9).   H

W. Schilz: Propagation of surface spin waves in the inhomogeneous field of rectangular YIG samples.
Phys. Stat. sol. (b) 53, 179-182, 1972 (No. 1).   H

W. Schilz: Experimental evidence for coupling between dipole and exchange dominated spin waves in epitaxial YIG.
Solid State Comm. 11, 615-616, 1972 (No. 5).   H

K. Schünemann (Valvo, Hamburg) & B. Schiek: Synchronization and noise performance of mutually coupled oscillators.
Arch. Elektronik & Übertr.technik (AEÜ) 26, 310-318, 1972 (No. 7/8).   H

E. C. Snelling: Some future aspects of magnetic materials and their applications.
IEE Coll. on New Materials, New Devices, New Techniques, Liverpool 1972, pp. 1/1-1/5.   M

J. L. Sommerdijk, W. L. Wanmaker & J. G. Verriet: Influence of host lattice on the infrared-excited visible luminescence in Yb^{3+},Er^{3+}-doped oxides.
J. Luminescence 5, 297-307, 1972 (No. 4).   E

S. Strijbos: Motion and distribution of large particles suspended in a fluidized bed.
Powder Technol. 6, 337-342, 1972 (No. 6).   E

T. J. B. Swanenburg: Improved geometry for a semiconductor surface-wave oscillator.
Electronics Letters 8, 351-352, 1972 (No. 14).   E

D. G. Taylor: The minimum brightness gain required in viewers using image intensifiers.
Optica Acta 19, 421-424, 1972 (No. 5).   M

A. Thayse: Multiple-fault detection in large logical networks.
Philips Res. Repts. 27, 583-602, 1972 (No. 6).   B

J. B. Theeten, H. Paradan, M. Escarieux, J. L. Domange (ENSCP, Paris) & J. Bonnerot (RTC La Radiotechnique, Suresnes, France): Appareillage pour étude par diffraction d'électrons lents à température variable dans la gamme 20 °K - 150 °K.
Rev. Physique appl. 7, 73-76, 1972 (No. 1).   L

J. P. Thiran: Polynomial recursive digital filters approximating the ideal low-pass characteristic.
Network theory, editor R. Boite, publ. Gordon & Breach, London 1972, pp. 519-537.   B

J. M. P. J. Verstegen, J. W. ter Vrugt & W. L. Wanmaker (Philips Lighting Division, Eindhoven): Luminescence of Eu^{2+}-activated $SrB_2Si_2O_8$.
J. inorg. nucl. Chem. 34, 3588-3589, 1972 (No. 11).

J. F. Verwey: Hole currents in thermally grown SiO_2.
J. appl. Phys. 43, 2273-2277, 1972 (No. 5).   E

W. L. Wanmaker, J. W. ter Vrugt & E. P. J. Meester (Philips Lighting Division, Eindhoven): Luminescence of alkaline earth zirconium silicates.
J. Electrochem. Soc. 119, 1237-1240, 1972 (No. 9).

G. Weimann, B. Elschner (both with II. Phys. Inst. Techn. Hochschule Darmstadt), K. H. J. Buschow & R. P. van Stapele: Paramagnetic resonance of gadolinium in the intermetallic compound $Gd_xY_{1-x}Ag$.
Solid State Comm. 11, 871-874, 1972 (No. 7).   E

J. S. C. Wessels & G. Voorn: Photochemical activities of chloroplast fragments obtained by the action of digitonin.
Proc. 2nd Int. Congress on Photosynthesis, Stresa 1971, pp. 833-845; 1972.   E

H. J. de Wit & C. Crevecoeur: The electrical conduction of glassy As_2Se_3 at high fields.
J. non-cryst. Solids 8-10, 787-792, 1972.   E

F. Wolf: Digital und analog anzeigendes Tera-Ohm-meter mit kontinuierlich einstellbarer Meßspannung.
Int. elektron. Rdsch. 26, 184-188, 1972 (No. 8).   H

J. M. Woodcock & J. E. Ralph: Electroluminescence from MIS and MIM structures.
J. non-cryst. Solids 11, 83-96, 1972 (No. 1).   M

The Netherlands astronomical satellite (ANS)

W. Bloemendal and C. Kramer

In the summer of 1974 an American space vehicle will put into orbit round the Earth a satellite that was designed and constructed (except for one of the measuring systems) in the Netherlands for astronomical research. The article below is the first of several that will appear in Philips Technical Review about the design and construction of this satellite (known in the Netherlands as 'ANS'), which has been entrusted to a consortium formed by Fokker-VFW and Philips. In this first article the authors have confined themselves to a general description of the configuration and operation of the satellite and to a brief account of the nature, history and organization of the project. Later articles will deal more thoroughly with some of the components of the satellite such as the attitude-control system, the onboard computer, and the sensors and actuators.

Nature and aims of the project

Work has been going on in the Netherlands for some time on a project whose aim is to put a small astronomical satellite into orbit around the Earth. The project is known in the Netherlands as 'ANS', an acronym for *Astronomische Nederlandse Satelliet*. This activity was preceded by an earlier phase that started in about 1965.

By this time it had begun to become clear that industry — not only in America, but in Europe too — received new stimuli ('spin-off') through taking part in advanced activities connected with space technology. In Europe a considerable amount of the work in space technology was carried out under the auspices of the European space agencies ESRO and ELDO, financed by the member countries, which included the Netherlands. Apart from these activities a number of European countries also had their own space programmes, which is still the case today. The effect of this was that the industrial organizations in these countries were well placed to serve as contractors for ESRO and ELDO, because of their greater experience, as was confirmed by the contracts that were given to these firms. In order to obtain the experience necessary for effective contribution to international space activities it therefore appeared that a national space programme would also be desirable for the Netherlands.

At about the same time it began to be apparent that space technology had much to offer astronomy in the

way of completely new possibilities: with instruments in satellites observations could be continued during long periods of time outside the undesirable influences of the Earth's atmosphere. There was considerable interest in these ideas in the Netherlands, a country that has long occupied a prominent place in astronomy. This could only partly have been satisfied by the limited facilities offered for installing observing instruments on board European or American satellites.

These incentives, which remain equally strong today, led to the formation of a working group, in consultation with the Ministries concerned, consisting of representatives from astronomical groups and industrial firms. There were representatives from the Departments of Astronomy at the Universities of Groningen, Leiden and Utrecht, from the industrial firms Fokker, Van der Heem and Philips, and at the initial phase also from the Netherlands Aero Space Laboratory (NLR). Two separate proposals emerged from the deliberations of the working group, both based on the same kind of scheme. These proposals were formally presented in July 1966; the astronomers presented their proposal to the Minister of Education and Sciences, and their industrial partners presented theirs to the Minister of Economic Affairs. In each of the proposals a plan for the technical realization of the project was included, accompanied by a brief time scale and a cost estimate. On the basis of these proposals and after positive advice from NLR the industrial firms were asked by the Netherlands Government to carry out a design study.

Drs W. Bloemendal is the Director of Space Activities of Fokker-VFW B.V., Amsterdam. Ir C. Kramer is with Philips Research Laboratories, Eindhoven. Drs Bloemendal is the Project Manager of the ANS project and Ir Kramer is the Chief Engineer.
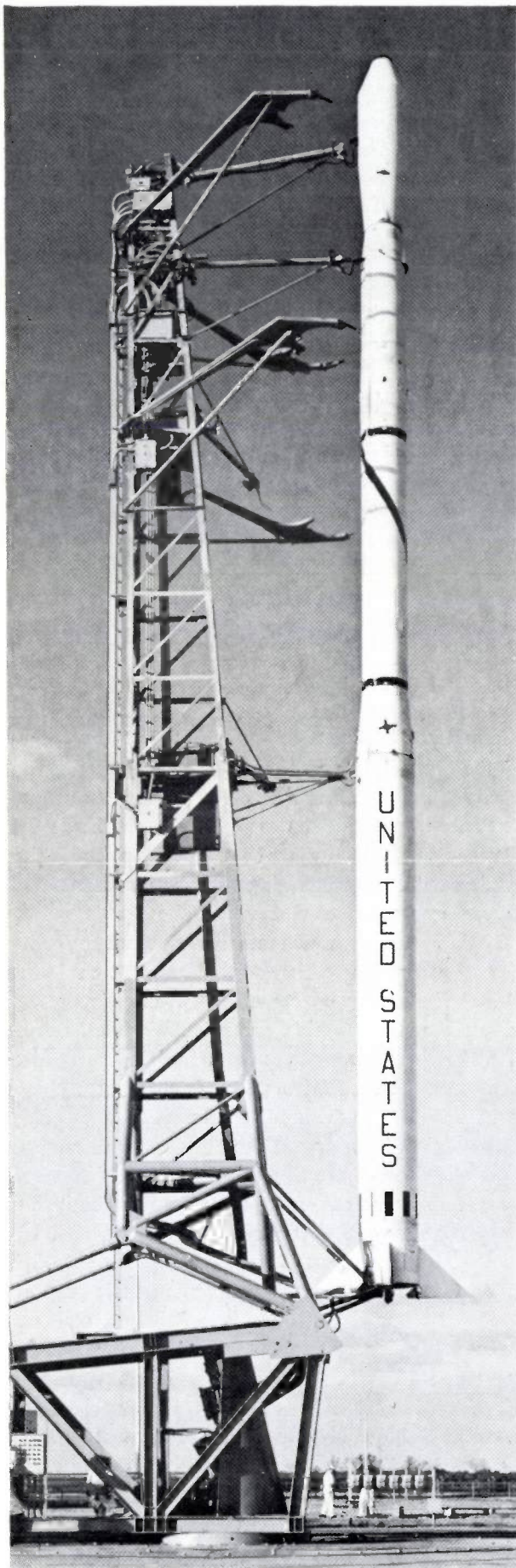
To enable this study — and later on the project, if one resulted — to be carried out in the most effective way consultation was set up between Fokker and Philips (Van der Heem had in the meantime joined the Philips group of companies). This consultation has led to the setting up of the 'ANS Industrial Consortium' (ICANS). This has the legal status of a partnership firm, with Philips and Fokker as the two partners. The board consists of four members: the chairman, the secretary and two executive members — the authors of this article. The agreement setting up the consortium was signed in November 1968.

In the meantime, in consultation with the Ministry of Economic Affairs, a start had been made in October 1968 with the design study, for which a contract was signed in January 1969. As the result of a number of detailed studies that had been made in the interim period, a number of modifications to the original proposal were included. The number of observation systems ('experiments') to be included in the satellite was fixed at two: one from the University of Groningen and one from the University of Utrecht.

As soon as the consortium had been set up an American advisory body was sought for the next phase of the project. After extensive consultation General Electric [*] was chosen, a company that is one of the leaders in American space technology.

The first activity of the study was to discuss the project thoroughly with the GE experts, who expressed their confidence in the feasibility of the project.

Since it was clear that any launch would have to be made with an American launch vehicle, contact was made in 1969 with NASA [**]. After a study of the satellite and its scientific objectives, NASA proposed that they should make available a Scout vehicle with increased thrust (*fig. 1*) and use part of the greater load-carrying capacity to include in the satellite an observation system of American origin. With this arrangement the ANS project obtained the status of a 'cooperative programme' with NASA; this was later confirmed by a 'memorandum of understanding' exchanged between the Netherlands Government and NASA.

The study phase continued until the end of 1969 and was completed in December by a detailed technical report [1], accompanied by a planning schedule and an estimate of the costs. On the basis of this report the Netherlands Government decided that it would instruct the consortium and the two scientific institutions to carry out the project under investigation. The contract for the project was signed in December 1970.

Fig. 1. A space vehicle of the Scout type on its launching tower. This four-stage rocket (length 22 m, launching weight 20 tons) will be used to put the Netherlands astronomical satellite (ANS) into orbit around the Earth.

The astrophysical observation systems

The central place among the observation systems to be carried by the satellite was taken by the equipment from the University of Groningen. It is desired to measure the brightness of a number of 'very blue' stars in the ultraviolet part of the spectrum, at five wavelengths situated between 1500 and 3300 Å. The objective in this work is to extend the existing classification of stars, which for young hot stars is rather uncertain, towards the ultraviolet. If such a classification is to be of genuine use measurement of a large number of objects is necessary. The limiting sensitivity of the equipment is therefore made equal to the tenth magnitude. Faint stars of this type are mainly to be found in the plane of the Milky Way. This requires that the equipment should only have a small field of view (2.5 × 2.5 minutes of arc) and should also be aimed ('pointed') with corresponding accuracy.

The instrument consists of a Cassegrain telescope, followed by a grating spectroscope, and contains five photomultiplier tubes, which act as detectors. The telescope plays an important subsidiary role in that it produces, via a system of lenses and mirrors, an image of the surroundings of the star being investigated on the photocathode of a star sensor. As will be explained later, this enables the reference of the attitude control to be produced by the instrument itself, so that alignment errors are largely eliminated. Since the full aperture of the telescope is used for this purpose, this is also the most sensitive way of obtaining the reference signal.

The instruments from the University of Utrecht will be used to measure the radiation from particular X-ray sources. The equipment consists of two parts: one for measurements in the wavelength range 44 to 55 Å, and the other for the ranges 2-4, 4-12 and 27-35 Å. This equipment can be used to detect faint X-ray sources; the spectral distribution and the time dependence of the strength of the radiation from these and other known objects can be measured. The measurement of time dependence — the investigation of the 'pulsars' — is a completely new field, in which remarkable results have already been achieved.

The American observation system to be mounted in the satellite is being constructed by American Science and Engineering (AS & E), who have designed it in close cooperation with the Massachusetts Institute of Technology (MIT). The objective is to perform measurements similar to those of the Utrecht group,

but in a different wavelength range (0.3-6 Å), and it will also be used to detect a pair of particular spectrum lines with the aid of a Bragg crystal. The existence of X-ray sources in the universe was first discovered by AS & E scientists, who are among the leaders in this field today.

The satellite

At a very early stage it had already been decided that the ANS would have to be launched by the smallest available vehicle, the Scout, to keep the costs of the project within reasonable limits. This implied immediately that severe limitations would have to be applied to the design concerning mass and dimensions. *Fig. 2* gives an impression of the exterior of the satellite; detailed drawings of the arrangement are shown in *fig. 3*.
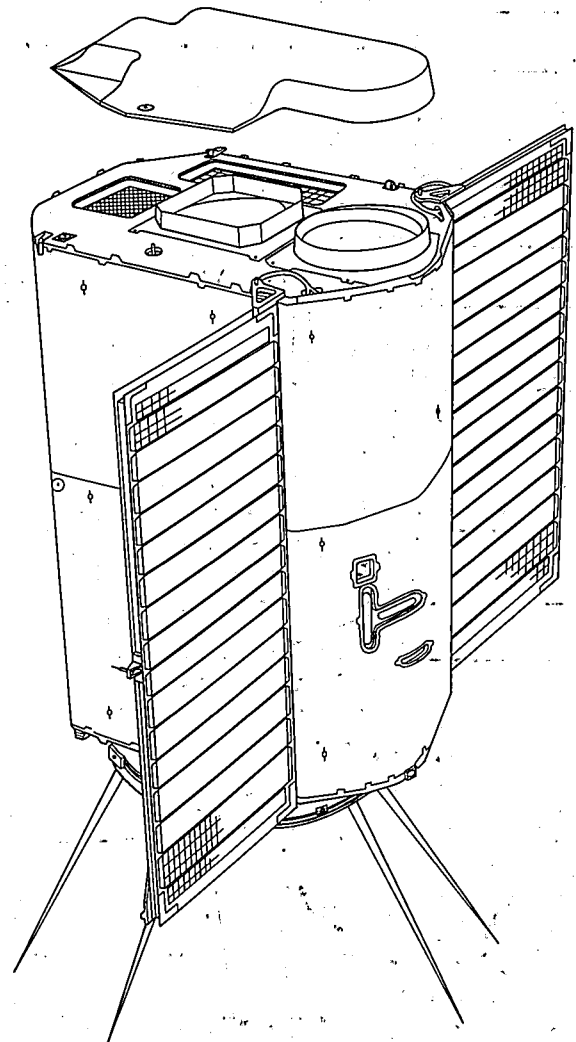
Fig. 2. The exterior of the satellite. The radiation to be measured reaches the measurement systems through the openings at the top. The solar panels can be seen on the left and the right in the deployed position. Between the solar panels there is an opening through which light can reach the solar sensors. The rods underneath are the aerials. The maximum total mass will be about 130 kg.

[*] General Electric Co., Philadelphia, Pa., U.S.A., Valley Forge Space Technology Center.

[**] National Aeronautics and Space Administration, Goddard Space Flight Center, U.S.A.

[1] A design study for an astronomical Netherlands satellite, performed by the ANS Industrial Consortium.
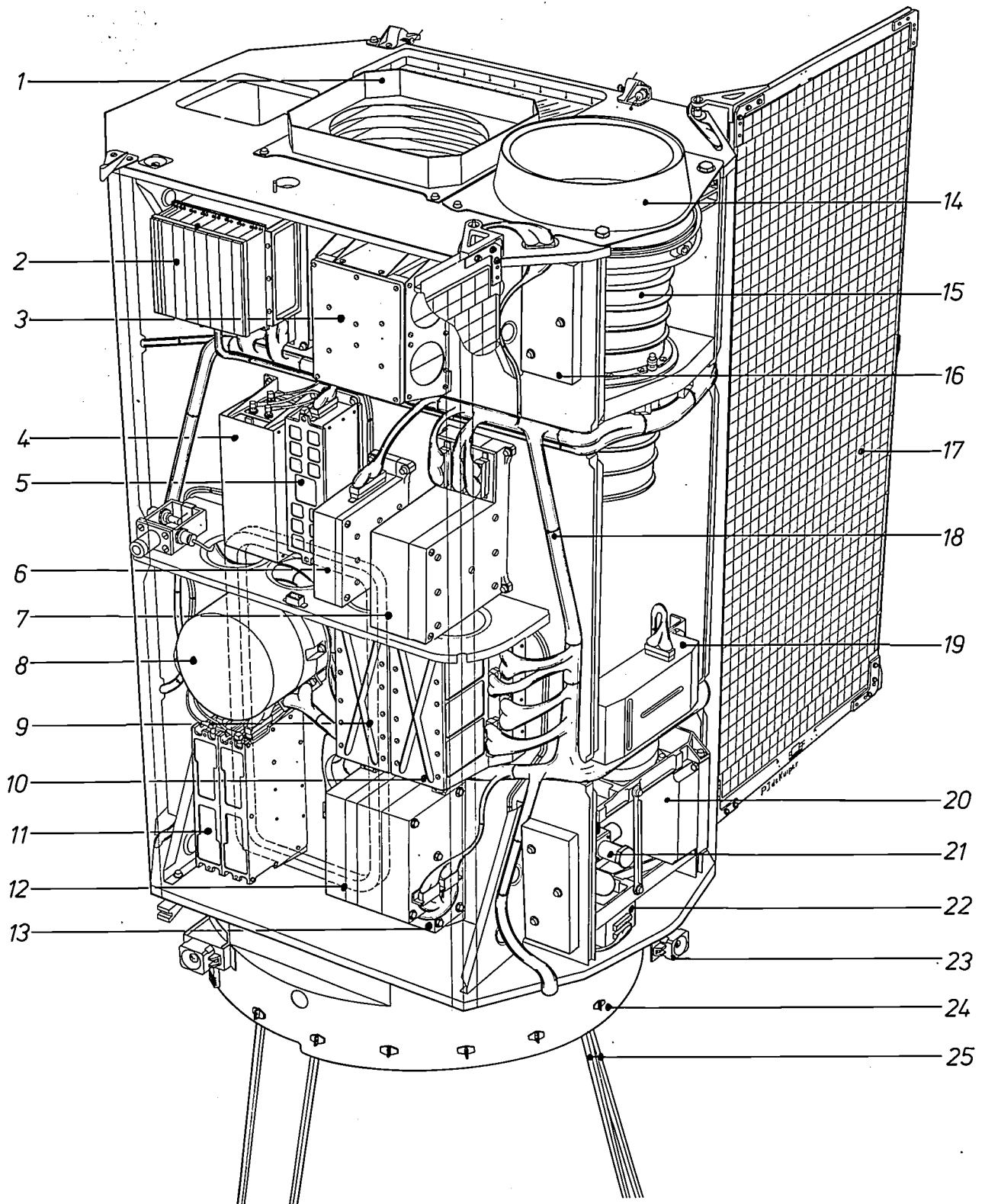
Fig. 3. *a)* Detailed arrangement of the Netherlands astronomical satellite (ANS). *1* shield to keep out scattered light. *2* power-supply control unit. *3* battery. *4* aerial coupling unit. *5* transmitter. *6* analog-to-digital converter for magnetometer signals. *7* electronic unit for attitude control. *8* reaction wheel for *y*-axis. *9* command decoder. *10* telemetry unit. *11* receivers. *12* magnet coil for *y*-direction. *13* data-processing circuit for soft X-ray radiation. *14* shield to keep out scattered light. *15* parabolic mirror, part of the observation system for soft X-ray radiation (44-55 Å). *16* balance weight. *17* solar panel (deployed). *18* cable-form. *19* solar sensors for *x*- and *y*-directions. *20* preamplifier for small X-ray detector. *21* gas-feed system for small X-ray detector. *22* electronic unit for gas-feed system. *23* coarse solar sensor. *24* 'yo-yo ring' (see the article of note [2]). *25* aerial.

Fig. 3. b) As for (a), but from the other side of the satellite. 1 parabolic mirror, part of the observation system for soft X-ray radiation (44-55 Å). 2 magnetometer. 3 voltage regulator. 4 voltage converter. 5 electronic unit for instrument for measuring ultraviolet radiation. 6 electronic unit of star sensor. 7 solar panel. 8 electronic unit for operating explosive mechanisms. 9 computer, consisting of seven memory blocks and a central data processor. 10 shield to keep out scattered light. 11 instrument for measuring hard X-ray radiation (2-12 Å). 12 preamplifier for large X-ray detector. 13 collimator with transmission channels grouped in honeycomb structure, part of instrument for soft X-ray radiation. 14 detector. 15 gas-feed system for large X-ray detector. 16 reaction wheel for z-axis. 17 horizon sensor. 18 balance weight. 19 yo-yo ring. 20 aerial.

The nature of the experiments, the observation of a large number of very faint stars, requires that the satellite should be accurately oriented and that it should be possible to alter its attitude easily. This is best done by three-axis stabilization, but the high directional accuracy required (1') is difficult to achieve in a satellite of low weight and therefore low moment of inertia. An acceptable solution that has been found to this problem is to make the satellite describe a polar orbit. For pointing the satellite, carrying out the measurement programmes and processing and storing the data obtained, it was found necessary to include a small computer in the satellite. Since this computer could only have a permitted weight of 16 mg per bit and a volume of 20 mm³ per bit, it required technology then in advance of existing developments. The required pointing accuracy obviously imposed very difficult requirements on the sensors and actuators that would have to be developed for the control functions.

The limited space in the nose cone of the rocket (*fig. 4*) and the spin stabilization of the last stage require deployable solar-energy panels, while a system is also required that will stop the rotation of the satellite after it has separated from the propelling vehicle.
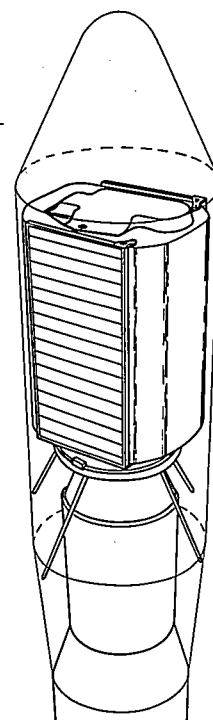
Fig. 4. The satellite, as it will be mounted on the fourth stage of the Scout vehicle in the nose cone. The solar panels are then folded in, and the openings at the top of the satellite (see fig. 2) are covered by an ejectable dust cover.
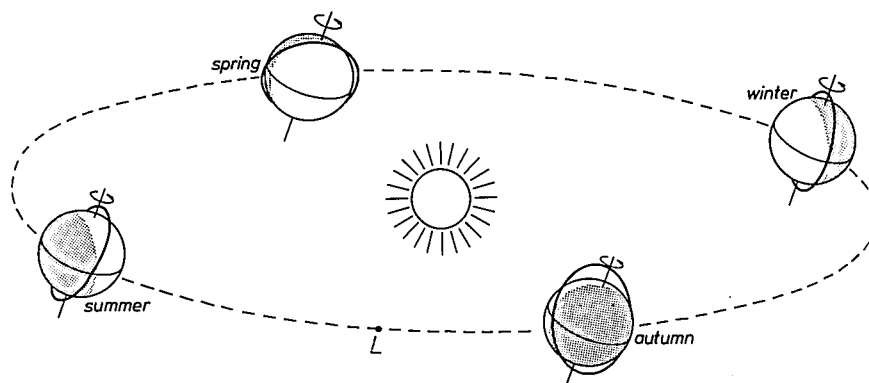
Fig. 5. The orbital plane of the satellite follows the movement of the Earth around the Sun, so that its normal always points towards the Sun. This keeps the satellite permanently in sunlight. The situation at the beginning of the four seasons is shown. The launch will be made in the summer (point *L*).

Finally, the vibrational forces that arise during launching are largest of all in small rockets. This imposed particularly difficult mechanical requirements on a small satellite of light construction that would have to carry delicate instruments.

Choice of orbit; attitude control

As has already been pointed out, the pointing of the observation equipment entails difficult requirements for the attitude control of the satellite. The Earth cannot be used as the stable platform, as in a terrestrial telescope, and the satellite must be pointed with the aid of optical references and using equipment for producing control torques [2].

The reference most easy to find from an orbit round the Earth is the Sun. An attitude-control system has therefore been chosen in which one axis is continuously directed towards the Sun. Then not only is there a clear attitude reference, but the solar cells are used in the optimum way. A disadvantage is that objects can only be observed if they are located in a plane perpendicular to the connecting line with the Sun. However, because of the annual rotation of the Earth around the Sun this plane rotates once per year, so that any object can be observed within a period of six months.

It is clear that the attitude-control system adopted can only operate if the satellite is in fact in sunlight and does not lie in the shadow of the Earth. By choosing an

orbit that lies above the poles of the Earth, it is possible initially to remain outside the cone of shadow of the Earth. If the orbit is exactly polar, however, the attitude of its orbital plane is invariant with respect to the fixed stars. Then, with the passage of time, the satellite would arrive in the Earth's shadow because of the orbital movement of the Earth around the Sun. An

chance of running into eclipse; and on the other hand it is desirable to make the height as small as possible to give the maximum launched weight and the least background interference from charged particles.

Once the satellite has been placed in orbit and aligned on the Sun, another reference object is necessary to enable the satellite to be rotated about the axis
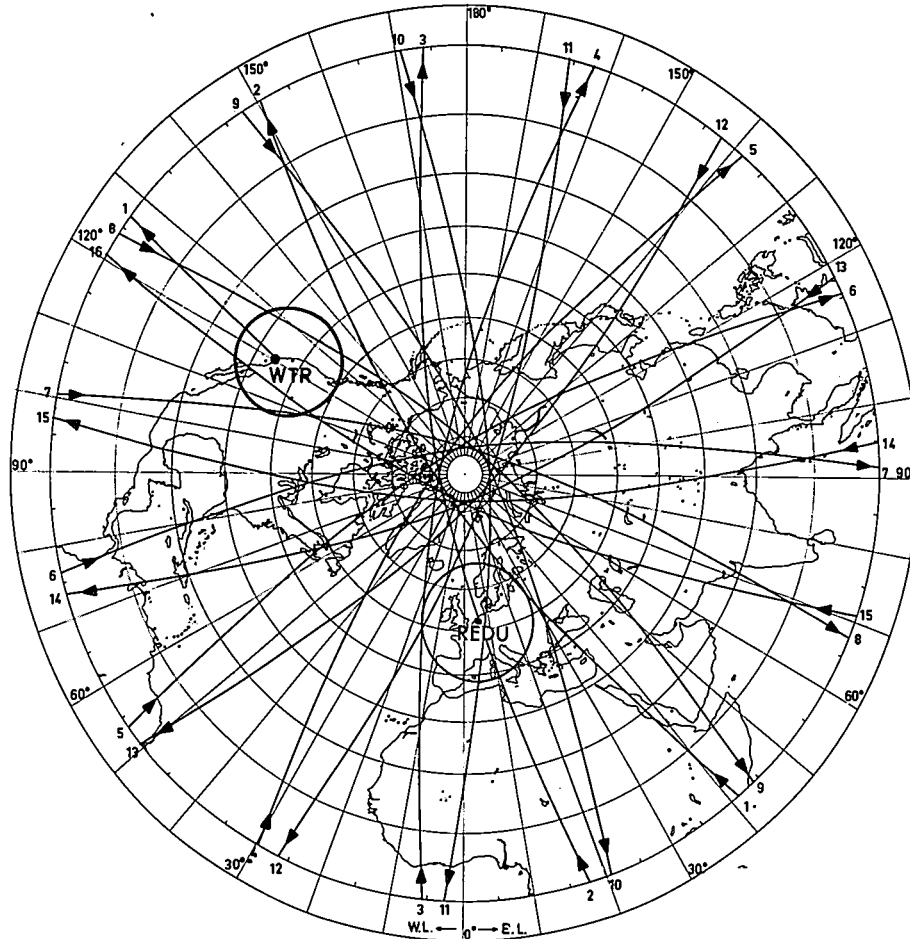
Fig. 6. Projection on the Earth's surface of the ground path of the satellite on the first day immediately after launching. As seen from the Earth, the orbital plane rotates with the Sun. WTR the Western Test Range launching site in California. REDU the ground station (see figs. 7 and 9). The circles indicate the region inside which radio contact is possible.

orbit is therefore chosen whose plane follows the movement of the Earth around the Sun (figs. 5 and 6). This is possible since the flattening of the Earth at the poles causes a certain amount of rotation of the plane of orbits that do not pass exactly over the poles. A suitable choice of the inclination of the orbital plane will give a situation in which the satellite remains outside the Earth's shadow for longer than six months at a time.

The height of the satellite in its orbit has been set at 500 kilometres. This value is a compromise. On one hand it is desirable to make the height as large as possible, to give as long an orbital life as possible, a large available time for ground contact, and a smaller

directed towards the Sun and into the correct attitude. This is done by making use of the infrared radiation from the Earth. A sensor with a small field of view scans the plane perpendicular to the direction of the Sun with a mirror rotating at 1 rev/s. In this way the angle is determined between the direction in which the telescopes are pointing and the two directions in which the sensor sees the horizon — 'rising' and 'falling'; these data are fed to the onboard computer.

The satellite can be pointed with an accuracy better than one degree with the horizon sensor. However, for

[2] A more detailed treatment of the attitude-control system is given in: P. van Otterloo, Attitude control for the Netherlands astronomical satellite (ANS); to appear shortly in Philips Technical Review.

pointing at stars this accuracy is not sufficient. A more accurate attitude control is not possible with a horizon sensor because the infrared horizon does not coincide with the optical horizon, but is located at a variable height in the atmosphere (30 ± 3 km), and also because the location of the satellite in its orbit is not known in advance sufficiently accurately for this purpose.

After the satellite has been pointed as well as possible with the horizon sensor, further attitude information is therefore obtained from a television camera tube of the image-dissector type located at the focal plane of the ultraviolet telescope. With this device the satellite can be pointed at a reference point, the 'guide-star', to the desired accuracy. An advantage of this method is that only those pointing errors arising in equipment directly related to the star sensor are of significance, whereas such effects as distortion in the telescope are largely eliminated because they have a similar effect on the guide-star and the object to be measured.

The attitude of the satellite is varied with the aid of the reaction torques of three reaction wheels mounted at right angles to one another and driven by motors. All the data from the attitude sensors for fine control are suppled to a computer, which calculates the torques required from these data and from an observation program stored in the computer [3]. Attitude control with the aid of a computer has the advantage of considerable flexibility. Not only can control laws be applied that would otherwise require the development of special circuits, but corrections can also be made at a very late stage, even during the flight itself, from the results presented.

To prevent information stored in the computer from being lost in the case of a temporary loss of power, the memory is of the core-store type.

Processing the data

Regular radio contact with the satellite is necessary both for supplying in turn the measurement program blocks and also for retrieving the results of the measurements. The time that elapses between successive contacts with ground control is determined by the orbit selected and the number of ground stations. Except in the immediate surroundings of the north or south poles, there is no single location on the Earth from which contact with the satellite would be possible during each orbit. However, since the poles are not the most suitable places for a ground station, in practice it is only possible to make contact from a single station at intervals of about twelve hours. This means that the satellite must be able to carry the measurement programs for a period of at least twelve hours as well as the results obtained in that period. *Fig. 7* gives a

schematic diagram of the way in which the information flows between the satellite and the observers on the ground.

The procedure is broadly as follows. For every period of twelve hours a program suited to the opportunities available to the equipment, the attitude and the position of the satellite in its orbit is made at Utrecht, with consultation between the three astronomical groups. This program is transmitted to the European Satellite Operation Centre (ESOC) at Darmstadt (*fig. 8*). This is the centre from which the satellite will be controlled, via a ground station located near the village of Redu in the south of Belgium (*fig. 9*). At Darmstadt the program is checked and coded so that it can be entered into the memory of the onboard computer, and real-time command instructions can be added. Transmission between Darmstadt and Redu is by cable data link, and the transmitter at Redu operates on a frequency of 148 MHz (wavelength about 2 m). When the satellite is passing over Redu it is possible to transmit instructions that have to be carried out by the satellite in real time in addition to the instructions for storage in the computer. The satellite transmits verification information back to the ground about the signals it has received.

While the satellite is out of sight of the ground station, its own transmitter sends out a 0.15 W signal at a frequency of 137 MHz. This signal, which can also be used for tracking the orbit, is modulated by a signal of 128 bits/second that provides data about the state of the satellite. At the same time the information coming from the observation systems and the attitude-control system is entered into the memory of the onboard computer. Six blocks of 4096 words of 16 bits are available for this function. A seventh block serves as an operating memory and also contains the measurement program. A cross-connection between the onboard computer and the telemetry equipment allows a small part of the 'housekeeping' information to be stored in the computer, and allows a small part of the results obtained to be transmitted. When the satellite comes into direct view of the ground station, the transmitting power is increased to 1.5 W and the information rate to 4096 bits per second. The information stored in the memories can then be transmitted to the ground. The housekeeping information and the command-instruction verification are interleaved with the memory read-out data. The quality of the signal is checked immediately at the ground station, so that the satellite

[3] A more detailed treatment of the onboard computer will be given in: G. J. A. Arink, The onboard computer for the Netherlands astronomical satellite; to appear shortly in Philips Technical Review.
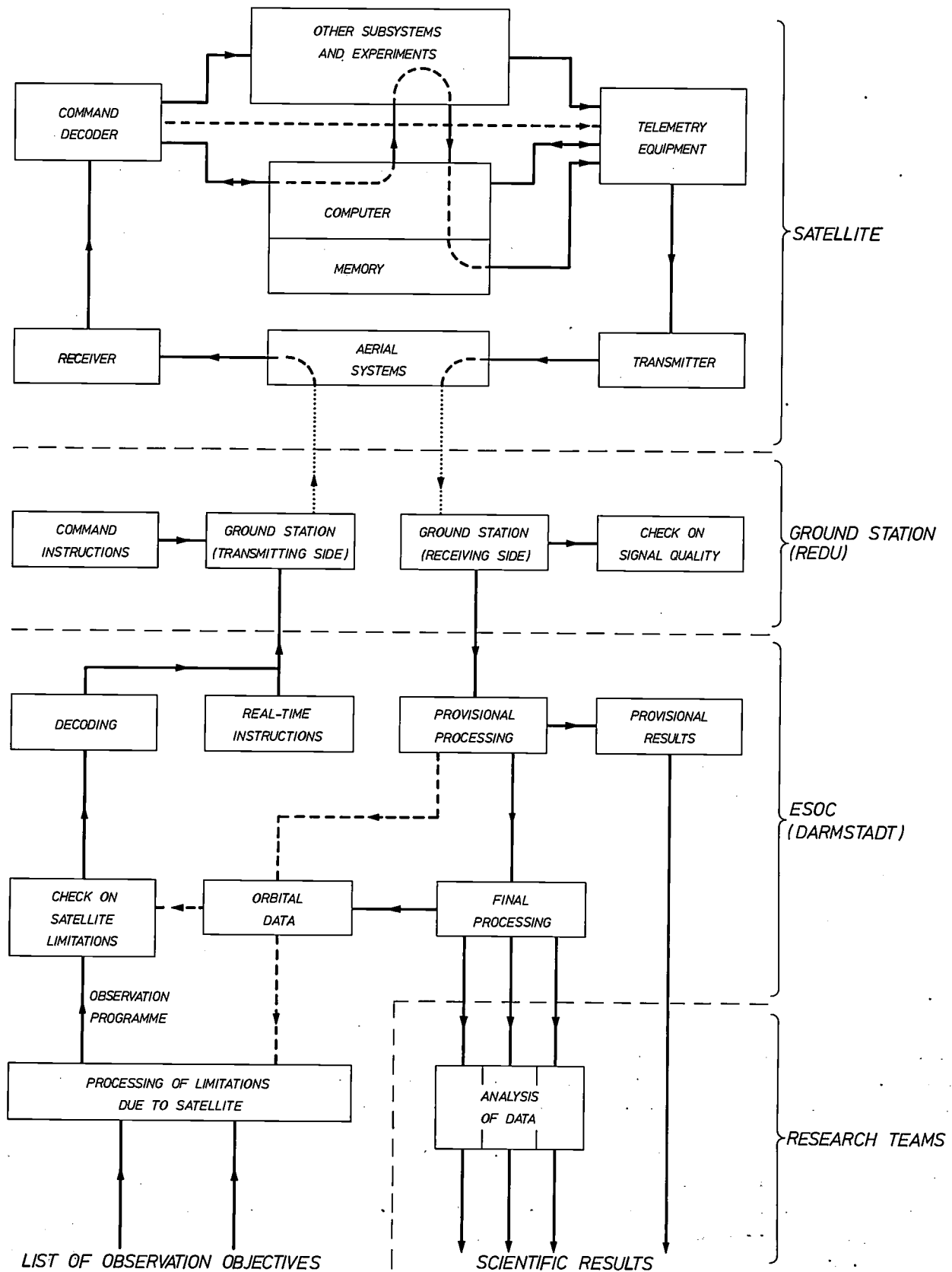
Fig. 7. Schematic representation of the way in which information flows between 'Earth' and the satellite. The programs to be carried out by the three observation systems in the next 12 hours are written at Utrecht and transmitted to the European Satellite Operation Centre (ESOC) at Darmstadt (West Germany). Here they are checked and coded, and the command instructions are added. The transmitter and receiver that maintain radio contact with the satellite are located at Redu (Belgium). Communication between Darmstadt and Redu is by cable.

Photo ESOC

Fig. 8. The control room of the European Satellite Operation Centre (ESOC) at Darmstadt (West Germany). The satellite will be controlled from here.

can be asked for example to retransmit a part of the information. The verification of the commands and the housekeeping information are processed during the ground contact. A provisional processing of the measurement data is then carried out, to obtain a rapid assessment before the next ground contact; this can be useful in deciding whether or not to transmit an alternative measurement program.

The final processing of the data is carried out after the exact orbital data have been obtained. The results obtained from this final processing are offered to the astronomers whose equipment is on board the satellite for scientific analysis.

Power supply and temperature control

The solar cells mounted on the deployable panels can produce between them a total power of more than 70 W. An NiCd battery (6 Ah) is also provided for use in periods when the panels cannot be used or can only be partially used (acquisition of Sun, aiming, periods of eclipse). The power-supply circuit provides two d.c. voltages, 20 V and 5 V; the 5 V supply is mainly used for digital integrated circuits.

The temperature control is passive, which means that the internal temperature of the satellite is determined by the equilibrium between the internal heat dissipation and the radiated heat. The thermal insulation is arranged in such a way that this radiation is emitted almost entirely by the upper and lower surfaces of the satellite. This is done to reduce as far as possible the temperature gradient in the transverse direction, which can cause mechanical distortion of the instruments that have to be aligned very accurately. For proper operation of the onboard systems the average internal temperature should lie between 0 and 30 °C. To keep the temperature within this range when some of the equipment has to be switched off extra heating elements are provided; these can be switched on from the ground.

The scheduling of the project

As was noted earlier, a start was made with the work of the project in early 1970 (*fig. 10*). In addition to setting up a sufficiently strong technical group a start was made with laying down the specifications and the
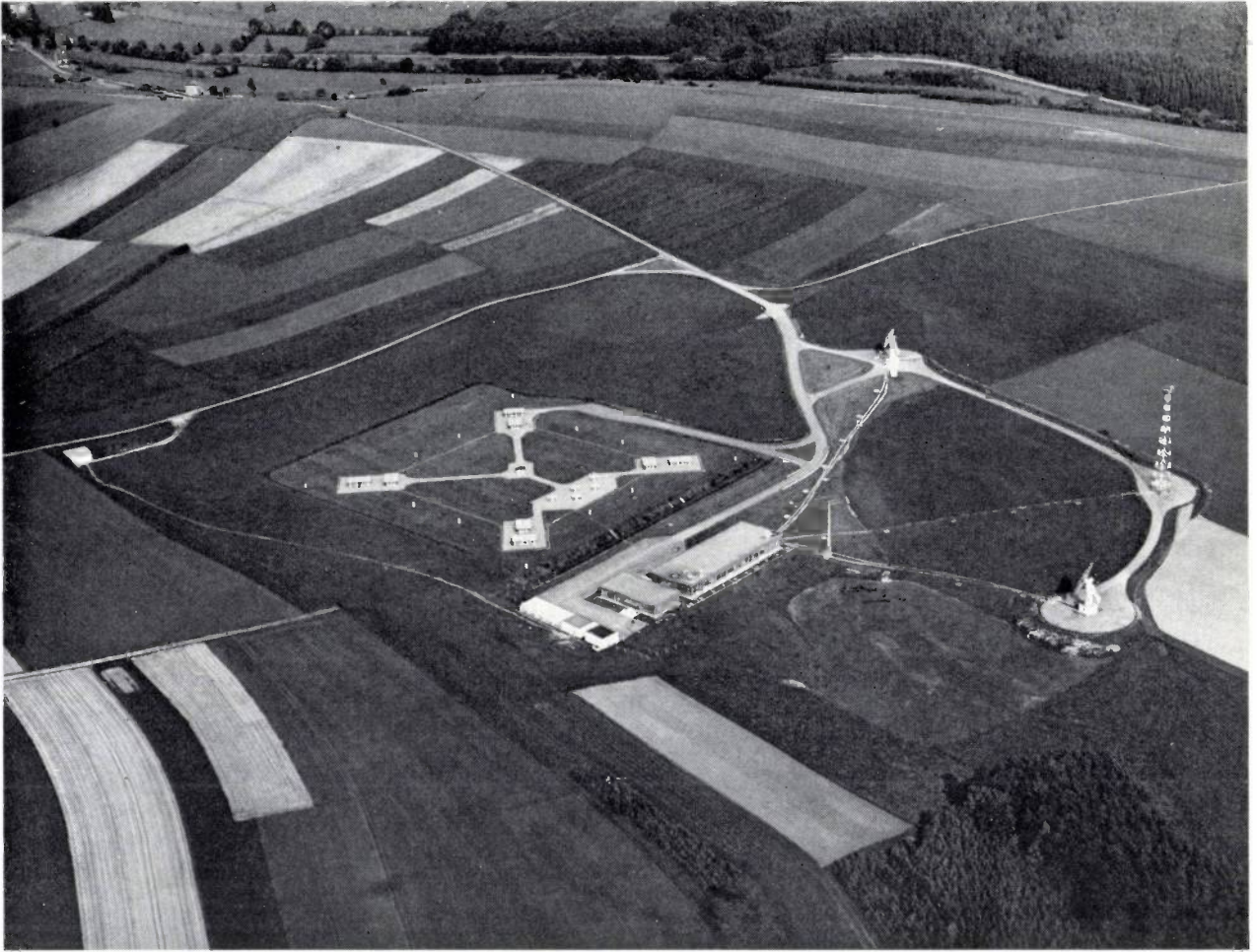
Fig. 9. The ground station at Redu (Belgium). ESOC will maintain radio contact of 10 minutes in every 12 hours via this station. The aerials with which the satellite will be tracked can be seen on the right.

development of the subsystems. The mechanical layout was checked with the aid of a dimensional model, with particular attention to the relative positions and ease of access of the various components. A 'prestructural' model was also made and used in vibration tests to check the main mechanical features of the satellite. Polar diagrams of the aerials were measured on a scale model and on a full-size mock-up.

At about the end of 1971 a start was made with the assembly — the 'integration' — of structural, thermal and electrical models of the satellite. The structural model was subjected to a static load test and to vibration tests, all intended to simulate the stresses that would be experienced on launching. The thermal model was mounted in a large vacuum chamber and exposed to simulated sunlight and to the low temperatures to be expected in space, to verify the thermal design of the satellite. The electrical model was first assembled on the bench and was then mounted in a satellite frame. In this way the interaction of the subsystems was checked in all possible operational situations.

At the start of 1972 a critical design review was carried out for each subsystem. After putting right inadequacies that had come to light in this review, a start was made on the construction of the equipment for the satellite proper. This will be integrated during 1973 and will put through an extensive test programme.

When this has all been completed the satellite will be sent to the launching site, checked there once again, and finally launched. After the launch the satellite will be checked and calibrated for one or two weeks in orbit. The phase in which the astronomical measurements are carried out then follows; this phase will have to be one of at least six months.

Organization

It is clear that good organization is necessary in the running of a development project of such complexity and with such a large number of participating groups

(*fig. 11*). The day-to-day management of the project is the responsibility of the Executive Coordination Committee (ECC). The two executive members of the ICANS board both have seats on this committee, as Project Manager and Chief Engineer, and so do the two Dutch Experiment Managers. The meetings are presided over by the Project Manager. The interests of the Netherlands Government in the project are represented by the Netherlands Agency for Aerospace Programmes (NIVR); the programme is supervised

responsibility of the Product Assurance Group, which again includes a representative from each company. The chairmen of these two groups report directly to the Project Manager, and the representatives to the Local Project Manager at their company.

It is of course impossible to include all the relationships within such a complex organization in a simple scheme. Activities such as 'integration' and 'operations' require some modification to the general approach, but we cannot enter into a detailed discussion here.
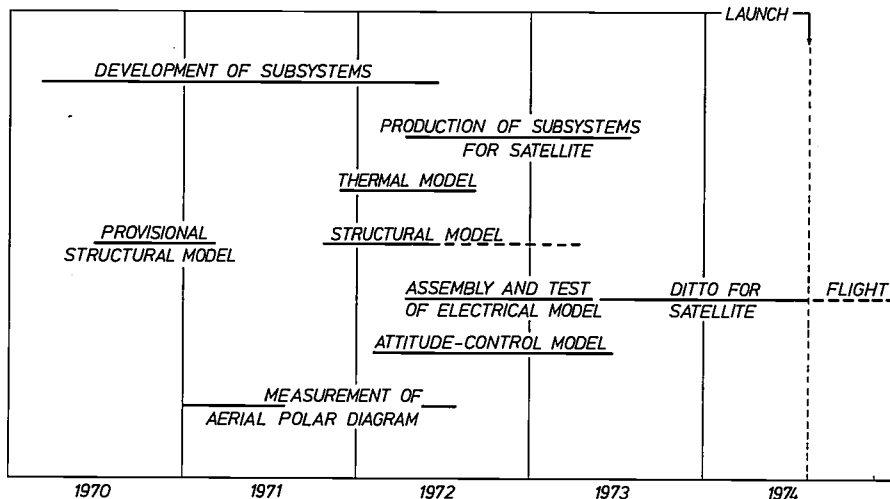
Fig. 10. Planning schedule for the complete programme of the project. The phase in which investigations are made on models and the subsystems are developed and constructed is now nearly completed.

by the ANS Supervising Director. Supervision of the scientific side of the project is the responsibility of a Project Scientist representing the Dutch astronomers. Both the Supervising Director and the Project Scientist attend the meetings of the ECC.

Within the ICANS organization the work is divided up into a number of activities that coincide in general with the subdivision of the satellite into subsystems. These activities are shared among the cooperating companies with each company producing one or more subsystems. In each company, one man — the Local Project Manager — is responsible for the coordination of all the activities for ANS within that company. The technical responsibility for this work is delegated to Subsystem Managers at the company.

The most important lines of communication are those from the Project Manager to the Local Project Managers for planning and financial accounting, and from the Chief Engineer to the Subsystem Managers for technical problems.

The progress of the project is monitored by the Project Control Group, on which each company has one of its planners as a representative. In a similar way, reliability and quality control of the satellite are the

Summary. In August 1974 a Scout space vehicle will be launched, from the Western Test Range launching site in California, which will put into orbit around the Earth a satellite made in the Netherlands and intended to carry out astronomical research. This Netherlands astronomical satellite (ANS) has been designed and is being constructed by the ANS Industrial Consortium (ICANS), a partnership between Fokker-VFW and Philips, at the request of the Netherlands Government. The satellite will carry three observation systems, from the Universities of Groningen (measurements on 'very blue' stars between 1500 and 3300 Å) and Utrecht (X-radiation from stars, 2-55 Å), and also from AS & E and MIT (Cambridge, Massachusetts; similar measurements, 0.3-6 Å). A polar orbit will be used such that the satellite will remain in continuous sunlight for at least six months. One of the axes of the satellite will be permanently directed at the Sun; searching for the objects to be measured will be done by rotating the satellite about this axis. Contact with the ground will only be possible once in 12 hours; this contact will be made from the European Satellite Operation Centre (ESOC) at Darmstadt via a station at Redu (Belgium). The instructions for the following 12 hours will then be fed into the memory of the onboard computer, which directs the operations from then onwards, and if required will send back the measured results at the next ground contact.

Fig. 11. The organization linking the various groups participating in the ANS project through the ANS Industrial Consortium (ICANS). The significance of the abbreviations is as follows:
FOKKER: Fokker-VFW B.V., Amsterdam.
PRL: Philips Research Laboratories, Eindhoven.
PTI: Philips Telecommunication Industries, Hilversum.
HSA: Hollandse Signaalapparaten B.V., Hengelo.
VDH: Van der Heem Electronics B.V., Voorburg.
NLR: The Netherlands Aerospace Laboratory.

A flexible method for automatic reading of handwritten numerals

M. Beun

II. The thinning procedure and determination of the special points

In the first part of this article [*] a broad outline was given of our method for reading handwritten numerals automatically. The development of a recognition scheme was dealt with in some detail, but only the principle of the thinning procedure was described and no indication was given at all of the way in which the special points are determined. In this second part of the article we shall take a closer look at these two subjects.

The thinning procedure

As described in part I, a skeleton is made from a matrix of occupied and unoccupied positions, the occupied positions indicating the numeral (see *fig. 13*). Thinning takes place in a number of cycles in which the edge points of the character are successively removed. In the first thinning cycle points on the circumference of the original pattern are removed, leaving behind a rather 'thinner' numeral. In the second cycle the edge points of this new pattern are removed, leaving the numeral thinner still. In this process of 'skeletonization' it is obviously necessary to make sure that nothing is taken away from the skeleton itself. Points are not therefore removed that would cause a break, nor are end points removed. The thinning procedure is continued until there are no points left that can be removed; the line pattern then remaining is the skeleton. Since only edge points are removed, a skeleton is obtained whose lines lie approximately in the middle of the 'thick' lines that form the numeral, so that the skeleton is a good representation of the basic shape of the numeral.

Before looking at the thinning process in more detail, it is first necessary to decide which points are to be regarded as edge points. In the 'checker-board matrix' used in our method, each position has four nearest neighbouring positions [11]; in *fig. 14* these are the positions *1, 3, 5* and *7*. We call these the 'orthogonal'

Ir M. Beun is with Philips Research Laboratories, Eindhoven.
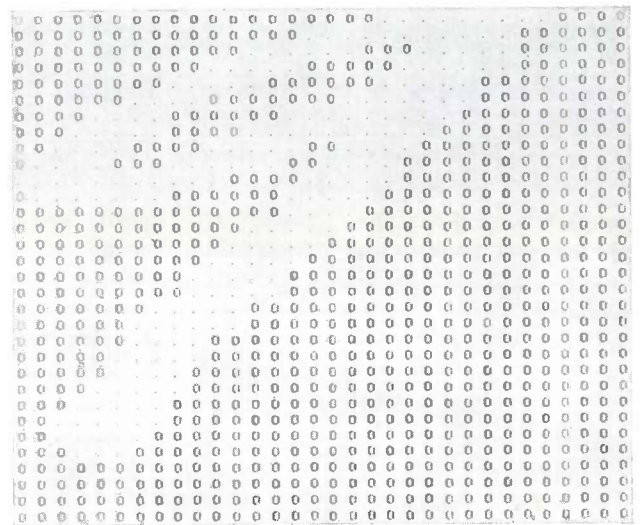
Fig. 13. Matrix with occupied positions (marked with a dot) and unoccupied positions (marked with a zero); the occupied positions represent the numeral *nine*. In this form the numerals are ready for skeletonization.

8	7	6
1	•	5
2	3	4

Fig. 14. Numbering of the neighbouring positions in the 'checker-board matrix' used in this article. The position marked with a dot has four 'orthogonal' neighbouring positions, numbered *1, 3, 5* and *7*, and four 'diagonal' neighbouring positions, numbered *2, 4, 6* and *8*.

neighbouring positions. We shall also consider positions *2, 4, 6* and *8* as neighbouring positions, and refer to them as 'diagonal' neighbouring positions. At this stage we shall regard a point as an edge point when it has one or more unoccupied orthogonal neighbouring positions; for the time being we shall not impose any constraints on the diagonal neighbouring positions, which may therefore all be occupied.

In part I we saw that a skeleton may also contain 'loop points'. These points must also not be removed,

of course, but this is of no significance during the skeletonization since loop points are never on the edge of a figure, and there is therefore no danger of them being removed if only the edge points defined above are removed.

Let us now return to the thinning process. The thinning cycles in which the edge points are removed each consist of two scans of the matrix, which we shall call the 'marking scan' and the 'stripping scan'. In the marking scan all the edge points of the numeral to be skeletonized are marked; as an example *fig. 15* shows the numeral of fig. 13 after the edge points have been marked with a letter *M*. The scan is in the usual sequence, from left to right, beginning at the top line. No points are removed during the marking scan, but only during the second ('stripping') scan. This scan only operates on marked points, each of which is checked to see whether it is a break point or an end point. If it is neither, the point is then removed — this is the stripping process. If a marked point turns out to be an end point or a break point, it is allowed to remain and the mark is cancelled.
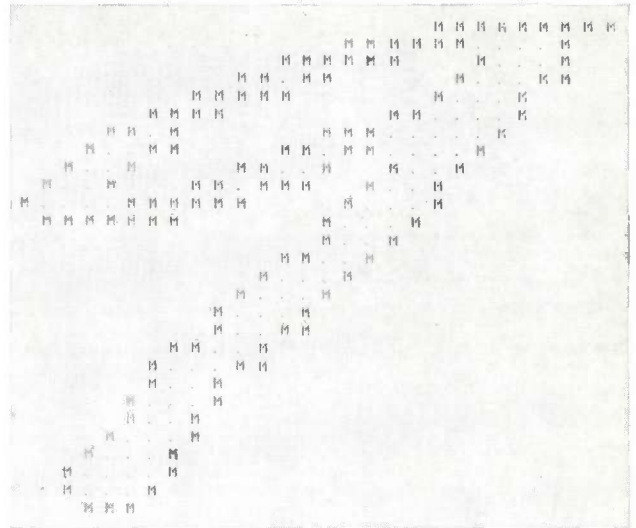
Fig. 15. The numeral in fig. 13 after the first marking scan; the edge points are marked with the letter *M* (the unoccupied matrix positions are left open here). An edge point is taken to be a point which has one or more unoccupied orthogonal neighbouring positions. The marking scan is followed by the 'stripping scan', in which the marked points are removed unless they are either break points or end points (loop points are not marked). This completes one thinning cycle.
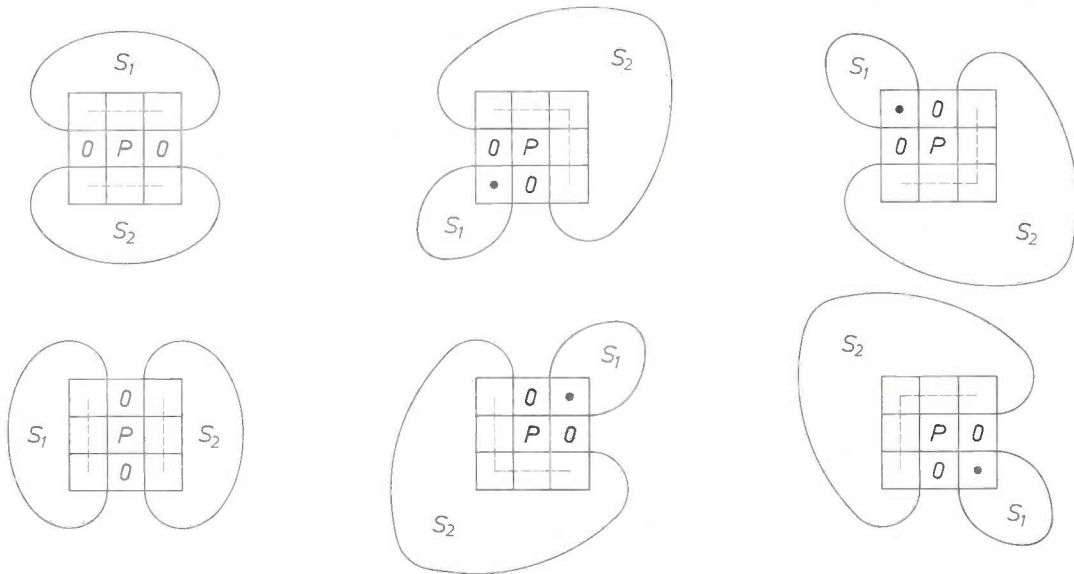
Fig. 16. The six possible configurations in which a point *P* is a break point. A dot denotes an occupied position and a 0 an unoccupied position; a broken line through a number of positions indicates that one or more of these positions is occupied. When point *P* is removed, these configurations divide into two parts S_1 and S_2. The two cases on the left are the same apart from a rotation through 90°, and the four cases on the right are the the same apart from rotations of 90°.

If the scan follows the *edge* of the figure to be thinned, then the edge points that are not break points or end points can be removed as they are encountered. No previous marking scan is then necessary. If there are loops in the figure, its outer and inner edges should be dealt with alternately to ensure that the skeletal lines run in the middle of the thick numeral. In this method the sequence in which the matrix positions have to be processed depends on the shape of the pattern to be thinned. This is no problem if a computer is used, but if the processing is to be done by special electronic circuits it is better to process the matrix positions in a fixed sequence. Thus in line-by-line scanning the process can be carried out in shift registers, as suggested by

P. Saraga and D. J. Woollons [12]. If however edge points that are not break or end points are removed directly in line-by-line scanning, the resulting skeleton is less satisfactory, since the skeletal lines do not run in the middle of the thick numeral, but are displaced to the right and downwards. This difficulty is avoided by introducing the marking scan.

[*]   M. Beun, A flexible method for automatic reading of handwritten numerals, I. General description of the recognition method, Philips tech. Rev. 33, 89-101, 1973 (No. 4).

[11]   Another type is the hexagonal matrix, in which each position has *six* directly neighbouring positions.

[12]   See the first article of note [4] in Part I.

During skeletonization the thinning cycle is repeated until a stripping scan is reached in which no further point is removed. The pattern then remaining consists entirely of break points and end points (and possibly some loop points), and thus forms the skeleton, or, to be more exact, not *the* skeleton but *a* skeleton. In fact, as will be shown below, there are still many ways in which the shape of the skeleton can be influenced.

During the stripping scans we must be able to determine whether a given point is an end point or a break point. End points are easily discovered; they are points with only one occupied neighbouring position (which may be either an orthogonal or a diagonal position). Break points are rather more difficult to trace; *fig. 16* shows the configurations in which a point P is a break point (in fact there are only two configurations, but they may occur in different orientations). An occupied position in this pattern is marked with a dot, and an unoccupied position with a 0, while a broken line through a number of positions indicates that one or more of these positions is occupied. We see that in all

Fig. 17. Skeleton produced from the figure *nine* in fig. 13 by several repetitions of the thinning cycle — consisting of a marking scan followed by a stripping scan. End points, ordinary skeleton points and fork points are indicated by *1*, *2* and *3* respectively. The skeleton is of no use for our recognition method because of the tails it shows at illogical places (there are five of them).

Fig. 18. The four edge criteria due to P. Saraga and D. J. Woollons. *a*) To determine whether a point P lies on the left-hand edge, the eight neighbouring positions of this point are divided into two groups A and B. Point P is considered to lie on the left-hand edge if the number of occupied neighbouring positions in region A is smaller than 2. and in region B greater than 2 ($n_A < 2$ and $n_B > 2$). The same requirement applies to the right-hand, upper and lower edges, except that the division of the neighbouring positions is then as shown in (*b*), (*c*) and (*d*) respectively. In our method we have combined these four criteria into one by treating a point as an edge point during the marking scan if it satisfies at least one of these edge criteria.

these cases the removal of the point P causes a break between the parts S_1 and S_2 of the pattern. If S_1 and S_2 are connected not only by P, but also at another place, the removal of P means breaking open a loop; in this case P is still a break point.

In the form described above, the thinning method does not always result in a good skeleton. In some cases, particularly with boldly written numerals (i.e. numerals that are thick in relation to the distance between two matrix points) 'tails' occur (see *fig. 17*), which make the skeleton less suitable for recognition. Avoiding tail formation is therefore an important problem in skeletonization. We have tackled the problem by introducing two modifications in the method described. Prior to the actual thinning, a 'smoothing procedure' is carried out; this 'irons out' protuberances in the contour that consist of a single position. As a result, fewer end points and therefore fewer tails occur during the thinning process. The principal change

made, however, consists in using a different definition of edge point for the marking scan from that given above. We shall now first discuss the new edge criterion and then show how it affects the skeletonization.

The Saraga-Woollons criterion

Our edge criterion for making skeletons without tails was derived from the work on optical character recognition by Saraga and Woollons at the affiliated British laboratories, Mullard Research Laboratories [13]. In the part of their work concerned with skeletonization [12] our British colleagues used separate criteria for the left-hand, right-hand, lower and upper edges. The criterion for the left-hand edge operates as follows. The eight neighbouring positions of the point under examination (P in *fig. 18a*) are divided into two groups, group A consisting of the neighbouring positions *8*, *1* and *2*, and group B consisting of the other neighbouring positions, i.e. *3*, *4*, *5*, *6* and *7*.

The numbers of occupied neighbouring positions in the two groups are called n_A and n_B respectively. Point P at the central position is now regarded as a point on the left-hand edge if $n_A < 2$, and at the same time $n_B > 2$. (The central point therefore has few points on its left, and a considerable number of points next to it on its right. Note that $n_A > 2$ means that the total number of occupied positions on the right of the double line is greater than *three*, since the central position is occupied by the point P.) The criteria for the other edges are the same as for the left-hand edge except that the areas A and B will be different; see fig. 18b, c and d.

We have combined the four criteria into a single one by treating those points as edge points that satisfy at least one of the four criteria. A consequence of using these criteria is that it is necessary to ensure that not only break and end points but also loop points are not removed during the stripping scan. This is necessary because some loop points are marked when the Saraga-Woollons criteria are used. *Fig. 19* gives two examples.
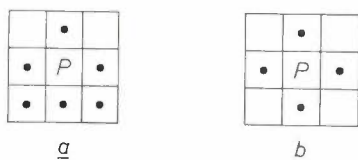
Fig. 19. Examples of loop points marked by the Saraga-Woollons criteria. Point P in (a) satisfies the criterion for the upper edge; point P in (b) manages to meet all four of the criteria. Since extra loops cannot of course be introduced during skeletonization, it is necessary to exclude not only the break and end points from removal during the stripping scan when using the Saraga-Woollons criteria, but also the loop points.

The central point in fig. 19a is a loop point that meets the Saraga-Woollons criterion for the upper edge (according to our original criterion it is not an edge point, and would not therefore be marked). The central point in fig. 19b is a loop point that manages to meet all four of the Saraga-Woollons criteria.

Using the combined Saraga-Woollons criterion for the marking scan is not in itself sufficient to produce better results. However, if for the stripping scan we also drop the requirement that no end points shall be removed — and retain this requirement only for the break points and the loop points — we then obtain good skeletons straight away without spurious tails. *Fig. 20* illustrates this with three skeletons all produced from the same numeral. For the first skeleton the simple edge criterion in fig. 15 was used during the marking scan; for the second skeleton the Saraga-Woollons criterion was used, but the prohibition on removing the end points was maintained; and for the third
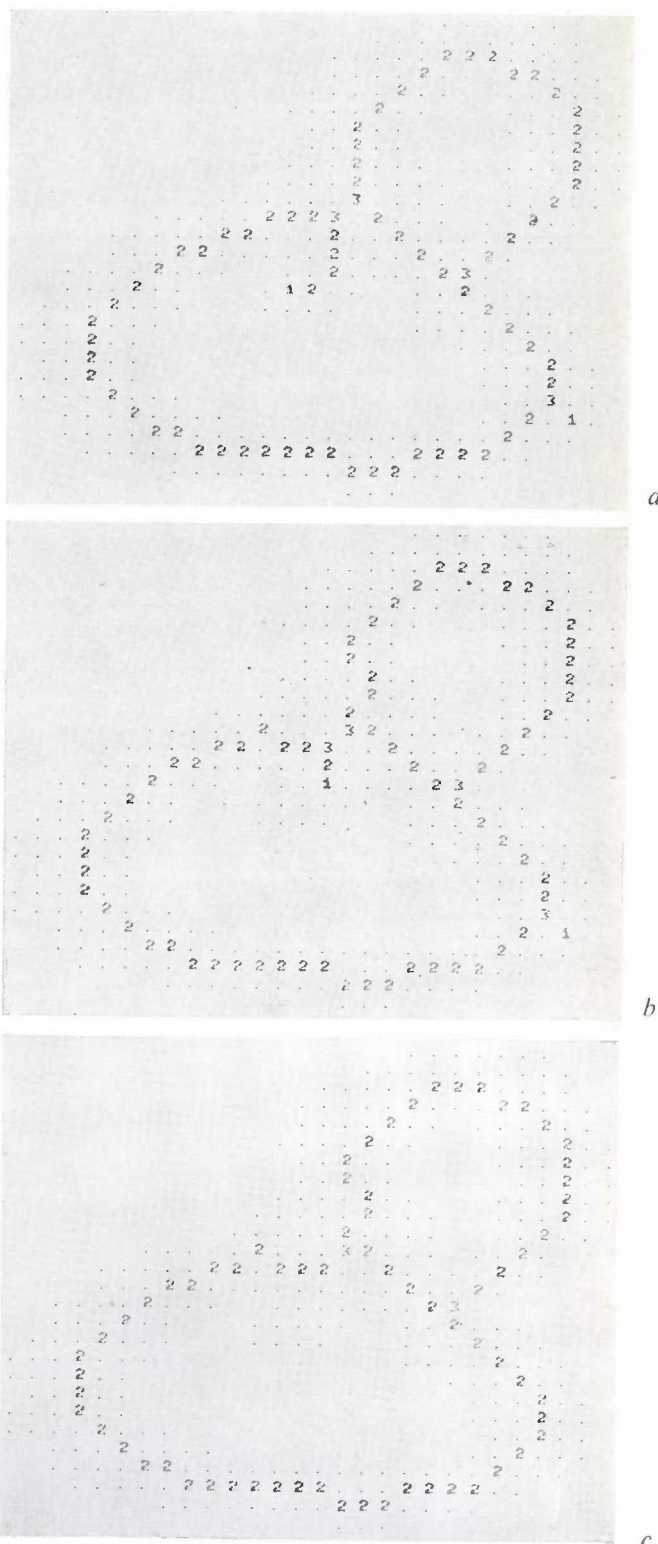
Fig. 20. Three skeletons of the same numeral *eight*, obtained with different thinning procedures. a) In mark scanning, the simple edge criterion of fig. 15 was used. The skeleton has two tails that interfere with recognition. b) In mark scanning, the Saraga-Woollons edge criterion was used, and the prohibition on the removal of end points was maintained in the stripping scan. Although the upper tail is shorter than in (a), both tails are still present. c) Here again the Saraga-Woollons criterion was used, but the prohibition on the removal of end points has been lifted. We see that no tails are formed in this case.

[13] See the first article of note [2] in Part I.

skeleton the Saraga-Woollons criterion was again used, but this time the prohibition was dropped, enabling end points to be removed during the stripping process.

At first sight the reason for this improvement of the skeletons is not clear. Points that satisfy the Saraga-Woollons criterion always have at least three neighbours anyway (because of the condition $n_B > 2$), and can therefore never be end points. This means that with the Saraga-Woollons criterion end points are never marked and thus cannot be removed in the stripping scan. The question whether or not end points may be removed therefore seems irrelevant here. If we take a closer look at what actually happens during the stripping scan, however, we see that a marked point that was not an end point during the marking scan can still

upper three rows can simply be removed (unoccupied positions are again marked with a 0). The first points in the bottom row do not present any special cases either, but if point m_1 is removed point m_2 becomes an end point, since it now has only one occupied neighbouring position, a diagonal one. The point m_2 is thus not removed, and its marking is cancelled (fig. 21c). In the marking scan of the second thinning cycle, six of the remaining seven points are marked, but not of course the end point that has just appeared (fig. 21d). In the subsequent stripping scan the first row again presents no special cases, but in the second row none of the points may be removed, since all of them are now break points. In the final state we are then left with an arm of four points (fig. 21e), which are left untouched by subsequent thinning cycles.
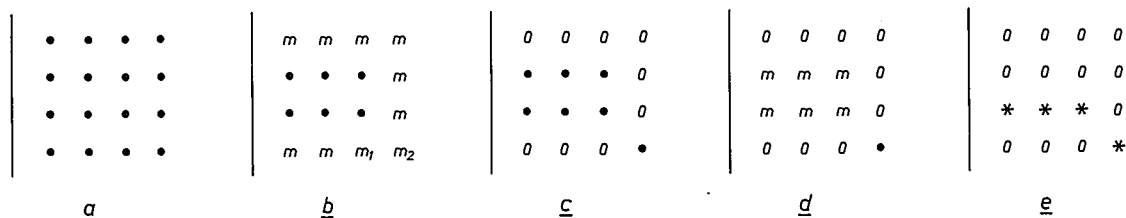
Fig. 21. Skeletonization of part of a numeral, using the Saraga-Woollons criterion and maintaining the prohibition on the removal of end points. a) Initial situation; the dots denote the occupied positions and the line indicates that there are more occupied positions on the left, which are not dealt with here. b) Situation after the first marking scan. c) Situation after the first stripping scan; unoccupied positions are marked with a 0. During the stripping scan, m_2 became an end point at the moment that m_1 was removed, so that m_2 can no longer be stripped. d) Situation after the marking scan of the second thinning cycle; the end point just formed is not marked. e) Final situation; the lower three marked points were break points and therefore had to be left during the stripping process. This configuration is no longer affected by further thinning cycles.
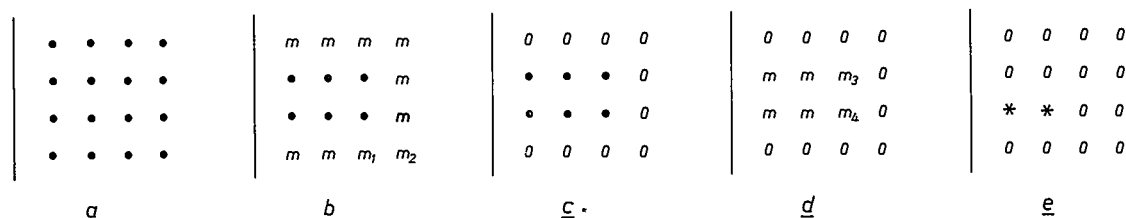
Fig. 22. The same situations as in fig. 21 but now during a thinning procedure in which the prohibition on the removal of end points has been lifted. The point m_2 can now be removed, and therefore so can the point m_4 during the second stripping scan, even though it became an end point at the moment that m_3 was removed. As a result, a much shorter arm is left in the final state than in fig. 21.

become an end point during the stripping scan before it reaches its turn to be stripped. Plainly, in this situation the removal of the prohibition does have a pronounced effect. This is illustrated in figs. 21 and 22.

Fig. 21a shows part of the matrix of a numeral to be skeletonized; the dots are the occupied positions and the line indicates that other parts of the character are situated on the left of the dots. We now follow the process during two thinning cycles in which the prohibition on the removal of end points is maintained. First the edge points are marked in accordance with the Saraga-Woollons criterion (fig. 21b). During the subsequent stripping scan the marked points in the

Fig. 22a to e gives the same 'instantaneous pictures' as fig. 21 but now during two thinning cycles in which the prohibition on the removal of end points has been dropped. We see that the end point m_2 is now removed (fig. 22c). Because of this, when point m_3 is removed in the second stripping scan (see fig. 22d), point m_4 becomes an end point. Nevertheless this point is also removed when its turn arrives during the stripping process, so that in the final state (fig. 22e) only two points are left over (these had to be left because they were both break points when their turn arrived during the stripping scan). We see that the arm remaining is shorter than in fig. 21, and that the figure

now gives a better reflection of the original character since it has lost the curvature at the end.

Excessive erosion

In the line scanning of the matrix there are two configurations of occupied points where there is a danger that complete arms of the original character will be 'eaten away' during skeletonization, causing the loss of essential information. *Fig. 23a* gives one example of this 'excessive erosion'. It shows a vertical, smooth bar with the thickness of two points, with an open end at the top. This situation may be present in the original character or may have arisen during the skeletonization. During the marking scan all points of the bar are marked as edge points in accordance with the Saraga-Woollons criterion. In the subsequent stripping scan

'Clean-up' scan

If the thinning process is continued, using the Saraga-Woollons criterion, until no further changes occur, it may nevertheless happen that the result still contains a few points that are neither end points, loop points nor break points. The point P in *fig. 24* is an example. It is not an edge point in terms of the Saraga-Woollons criterion; because of the condition $n_B > 2$, a point of this type always has at least three neighbours, and P has only two. In subsequent thinning cycles P will consequently not be marked and will therefore not be removed, so that when the thinning procedure finishes P will still be present. To remove any such points the thinning process is always followed by a single 'clean-up' scan. This is a scan that removes all points that are neither end points, break points nor

Fig. 23. *a*) Configuration of occupied points threatened by 'excessive erosion'. In accordance with the Saraga-Woollons criterion, all the points of the vertical bar are marked, and if no protective measures are taken, they will all be removed during the next stripping scan because they will never be break points or loop points. *b*) To avoid excessive erosion, we check for each point P that is removed whether only the neighbouring positions *3*, *4* and *5* are occupied; if this is the case, any marking of the point at position *5* is cancelled.

Fig. 24. Example of a point that does not belong to the skeleton, since it is not an end, break or loop point, but nevertheless is not removed in the thinning procedure. Point P does not satisfy the Saraga-Woollons criterion, which means that it is not marked and therefore not stripped. To avoid such points, we end the thinning procedure with a 'clean-up' scan, which removes all points that are not end, break or loop points.

all these points would be removed in turn, since at the moment when it is about to be removed none of these points is a break point or a loop point.

To avoid this process of erosion a check is made every time a point is removed during the stripping process to find out whether the neighbouring positions of the point are occupied in the way shown in fig. 23*b*. If this is found to be the case, any marking of the point at position *5* is cancelled, so that during the rest of the stripping scan this point cannot be removed. In the application of this procedure the marked points in the left-hand column of the bar are all removed, but the points on the right of the column are retained. In a subsequent thinning cycle none of these points will then satisfy the Saraga-Woollons criterion, so that the original configuration is represented in the skeleton by a bar with a thickness of one point.

The second configuration in which excessive erosion can occur is a horizontal bar with a thickness of two points, which is open-ended on the left-hand side. It is easy to show that the approach described above is also adequate in this situation, although the residual bar is then slightly bent at the end.

loop points when their turn arrives, and therefore it removes P. The clean-up scan is thus *not* preceded by a marking scan.

The use of a clean-up scan enables the number of thinning cycles to be limited in advance, e.g. to three. Even when the stage at which thinning causes no further change has still not been reached after these three cycles, a single clean-up scan produces a result consisting exclusively of end, break and loop points, i.e. a skeleton. This skeleton will not usually be very different from the result of skeletonization carried right through to the end.

Our thinning method works very well with the Saraga-Woollons criterion and we have therefore used it in the experiments described in part I of this article. A more detailed analysis of the thinning process in our later investigations has shown however that it is possible to find other edge criteria that will give good skeletons. We shall not pursue this further here [14].

[14] A forthcoming article in Philips Research Reports will deal in greater detail with some fundamental aspects of the skeletonization process.

The special points

In tracing the special points of the skeleton required for our character-recognition method — the end points and the fork points — every skeletal point is examined to determine how many arms go out from it. If we imagine the matrix positions occupied by a skeletal point to be black and the others white, then the arms extending from a point will be separated from each other by as many white areas as there are arms. We now determine the number of arms by counting the

that the conclusion based on the black-to-white transitions is correct in this case. A similar difference is found with the right-hand neighbouring point in fig. 25c and with all four neighbouring points in fig. 25e.

Fig. 26b presents the same configuration as in fig. 26a, but now the diagonal-neighbour connection is made visible by a line. This figure makes it clear that what we really have is not a fork point but a 'branching
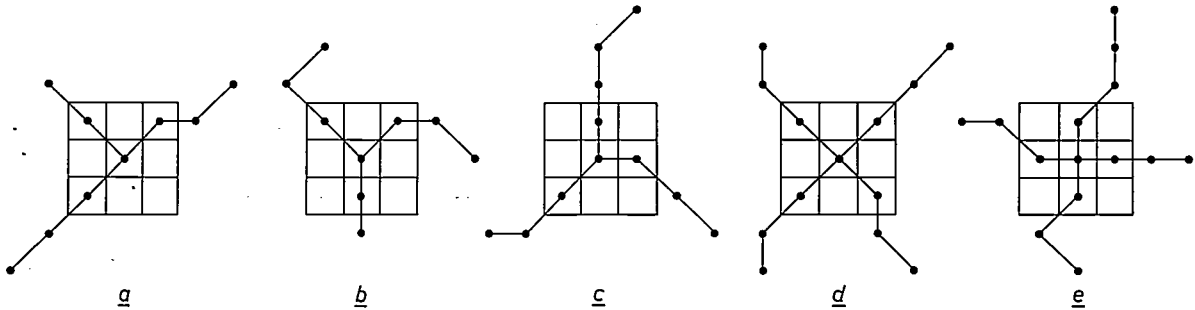
Fig. 25. Examples of fork points. If we imagine the matrix positions occupied by a skeleton point to be black, and the other positions white, the fork points can be detected by counting the number of black-to-white transitions in a search of the eight neighbouring positions of the point; this number is equal to the number of arms going out from that point. In the examples given here, the same result would be obtained by counting the number of occupied neighbouring positions of the point, but this method leads to errors in some other configurations.

number of black-to-white transitions in a complete search along the eight neighbouring positions of the point in question (we can of course also count the number of white-to-black transitions) [15]. If this number is one, then the point is an end point; if there are two transitions, it is an 'ordinary' point. Three transitions give a T-junction and four give a four-junction. Four is the maximum, because the eight neighbouring positions are then alternately black and white. According to the definition of a fork point given in part I, every T-junction is a fork point, whereas a four-junction must be treated as two T-junctions.

The simple method of determining fork points described here nearly always yields the right result; some examples are given in fig. 25. In all these cases the more simple-minded method of taking the number of arms as being equal to the number of neighbouring positions occupied by a skeletal point would also give the correct result. For certain other points, however, this method fails. We can illustrate this by considering the neighbouring position above the central point in fig. 25c; in fig. 26a this point is taken as the central point. In a search of the neighbouring positions here we count two black-to-white transitions (conclusion: no fork point), but the point has three occupied neighbouring positions (conclusion: there must be a fork point after all). It is evident

cluster' consisting of three points, from which three arms or branches extend. A cluster like this is much more difficult to handle than a simple fork point. With this simple cluster the method of black-to-white transitions automatically selects one of the three points as the fork point (i.e. the central point in fig. 25c) and classifies the two others as ordinary points. However, this method also fails when the branching cluster consists of more than three points, as can be seen from the example in fig. 27. The centre of the configuration in this figure is formed by a cluster of four points with four arms going out from it. A cluster of this kind may
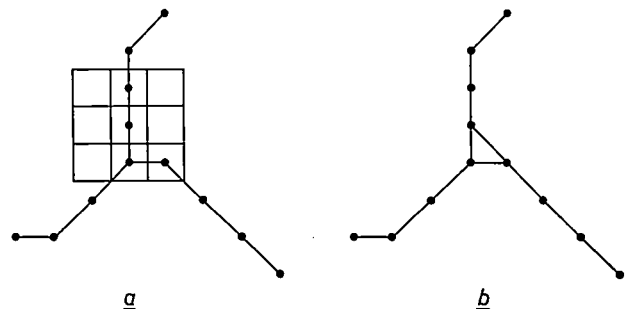
Fig. 26. a) Configuration in which the central point would wrongly be classified as a fork point if the number of neighbouring positions occupied by a skeleton point were to be used as a criterion. It is correctly classified as an 'ordinary' point (it is in fact the point above the central position in fig. 25c) by counting the number of black-to-white transitions in a search of the neighbouring positions. b) The same configuration as in (a); the line joining the three middle points shows that they form a 'branching cluster'. When the black-to-white transitions are counted only one point of this cluster is classified as a fork point.

remain after the thinning process, since the four points are all break points. Each of the points has *two* black-to-white transitions in its group of neighbours; in the search for fork points they are all therefore classified as ordinary points, which means that the cluster is not detected.

Theoretically it is possible to construct even larger branching clusters, which are also not affected by our thinning procedure [16], but so far we have not found any in skeletons of numerals. Even a cluster of four points, as shown in fig. 27, is extremely rare. For these reasons we have used the method of black-to-white transitions in our system of searching for fork points, but we have extended the method with a procedure for detecting four-point clusters. If such a cluster is found, the points at the lower left and upper right are then regarded as fork points, and the two others as ordinary points.

Fig. 27. Configuration showing a branching cluster consisting of four points. The method of counting the black-to-white transitions also fails in this case; for each of the four points of the cluster we only find two transitions, so that all these points would be classified as ordinary points. This error can be avoided by detecting the four-point clusters separately, and then treating two of the points as fork points and two as ordinary points.

Summary. This second part of the article explains in detail the methods for skeletonizing a numeral and for identifying the end points and fork points in the skeleton that are needed for recognition of the numeral. The skeletonization takes place in a number of cycles; in each thinning cycle the matrix with the occupied and unoccupied positions that represent the numeral are scanned twice. During the first scan the edge points of the numeral are marked (marking scan) and during the second scan the marked points are removed if they are not break or end points (stripping scan). Various measures are required to ensure that the skeleton does not have 'tails' that have nothing to do with the basic shape of the numeral. The choice of the criterion by which a point is declared to be an edge point is very important. Good skeletons are obtained with an edge criterion used by Saraga and Woollons at Mullard Research Laboratories. The fork points are easily traced by counting the number of black-to-white transitions in a search of the eight neighbouring positions for each point. If this number is one, two, three or four then the point is an end point, an ordinary point, a T-junction or a four-junction respectively. This method presents problems only if clusters of occupied points remain after the stripping scan. Such clusters, which are rare in skeletons of numerals, have to be separately detected and can then be split into ordinary fork points.

[15] The method of counting black-to-white transitions to determine the number of arms was proposed by P. Reijnierse of Philips Research Laboratories.
[16] This point will also be dealt with in the forthcoming article (see note [14]).

Heat pipes

G. A. A. Asselman and D. B. Green

II. Applications

The applications of heat-pipe systems that will be dealt with in this part of the article [*] are discussed under four distinct headings corresponding to four of the main features of heat pipes. These four features are: the exceptionally high thermal conductivity, the small variation in temperature throughout the pipe, the ability to give a heat-flux transformation, and the temperature sensitivity. The heat flux is transformed if the evaporator and condenser have different surface areas, so that the heat flux at these two parts of the pipe has a different value. Temperature sensitivity refers to the degree to which the mean temperature of a heat pipe depends on the heat flux, which is also a quantity of considerable significance in applications of the heat pipe. The temperature sensitivity of simple heat pipes is already very high, and it can be increased to exceptionally high values in the variable-conductance heat pipes, which we shall discuss in the concluding section of this article.

The four features mentioned above are of course interdependent and are used in combination in any single application. The justification for discussing the applications in groups, in each of which one of the four characteristic features is prominent, is that it may heighten the appreciation of the peculiar advantages of heat pipes over conventional heat-transport systems.

Heat transport

There are numerous applications employing the high heat-transfer capability of heat pipes. One that we have been concerned with is the transport of heat from a source of thermal energy to the heater head of a Stirling engine. When a heat pipe is used a Stirling engine can be combined with various types of heat source [7].

Fig. 7 shows a 10 hp Stirling engine being driven on a test bench by heat supplied along a 15 cm diameter sodium-charged heat pipe, operating at about 750 °C. In this instance the heat pipe received its heat input from a series of electrical heating elements (25 kW) wound around the evaporator section. In practice a

Stirling engine will not be required to operate from an electrical source, but from a combustion process. The advantage of incorporating heat pipes then lies more with the ability to transform heat fluxes than with the high heat-transfer capability. With unconventional combustion processes both aspects can play an important role. In addition, some Stirling systems currently under development are only viable because of heat-pipe technology.

Another application is a siphon for the transport of molten salts. As *fig. 8* shows, one end of the siphon is placed in the vessel containing the molten salt and heat is transferred from the salt and along the length of the siphon by heat-pipe action. In this way the siphon tube may be maintained at a temperature close to that of the molten salt in the vessel, enabling the salt to be transferred through the siphon without the danger of
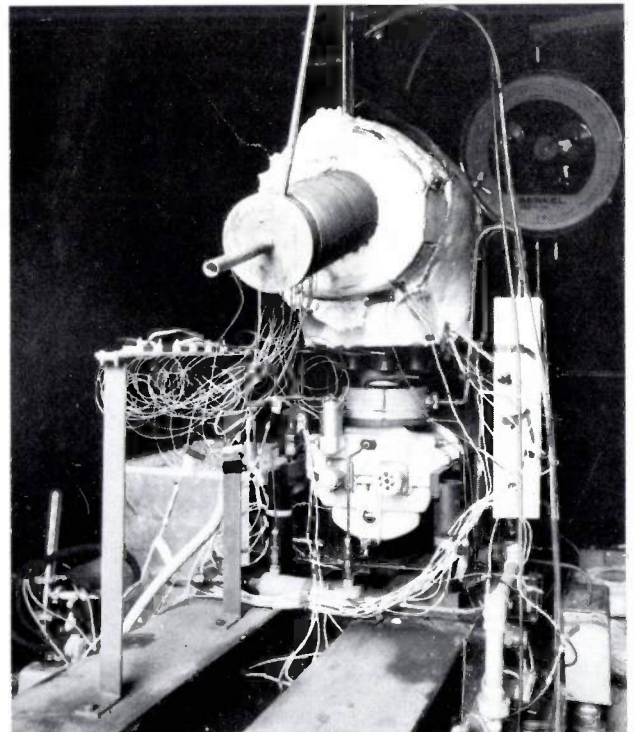
Fig. 7. Test system with a 10 hp Stirling engine driven by heat supplied along a sodium-charged heat pipe. The pipe itself is electrically heated; at 750 °C and with a temperature difference of only a few degrees it can transmit more than 25 kW of energy. The crankshaft of the Stirling engine can be seen at the bottom of the picture in the centre; above it is the engine head, enveloped in insulating material. The heat pipe protudes from the front.

Ir G. A. A. Asselman is with Philips Research Laboratories, Eindhoven; Dr D. B. Green was formerly with Philips Research Laboratories.
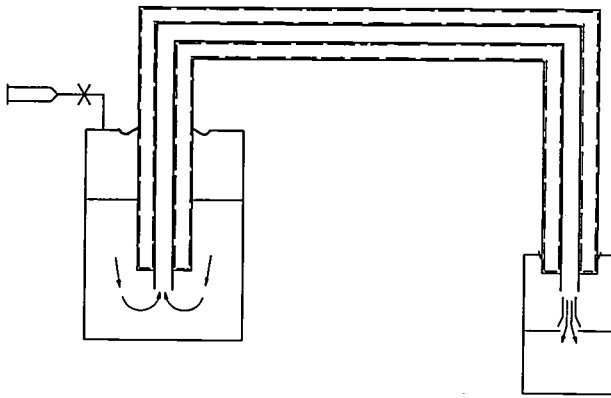
Fig. 8. A siphon heated by heat-pipe action. A siphon of this type can be used for transporting molten salts from vessel to vessel without the danger of clogging because of solidification. The evaporator section (on the left) is immersed in the liquid to be transported, so that the siphon tube is automatically kept at about the same temperature as the molten salt. The liquid flow is caused by the pressure of compressed gas admitted to the reservoir vessel. The insulating material around the pipe is not shown.

clogging arising from solidification. We have made a siphon of this type for the transport of molten lithium fluoride. The working fluid is sodium and the envelope is made of stainless steel.

Elsewhere the large heat-transfer capability of heat pipes has been applied in quite different ways. Some attempts have been made to cool electronic circuits and components with heat-pipe systems [8], thus eliminating the need for metal cooling bars and other heat sinks. Extensive use has also been made of heat pipes in direct-energy-conversion systems [9].

Isothermal spaces

The temperature is almost the same at all points inside a heat pipe. This feature arises because the thermal resistances R_{L-v} and R_{V-L} (see I, fig. 5) are small, so that the temperature gradients are also small, provided that the heat fluxes are not too excessive. A heat pipe whose basic function is the creation of an isothermal surface rather than the transport of large quantities of heat is usually referred to as an isothermal chamber. Such chambers are designed to have a minimum of obstruction to the vapour flow, so that pressure gradients in the vapour may be rapidly eliminated. Inside an ideal isothermal chamber the liquid in the wick is everywhere in equilibrium with its vapour. As we have seen (I, page 108/9), it follows from the relationship between the temperature and the pressure of the saturated vapour that a small change in temperature of the liquid in the wick can give rise to vapour flow towards or away from other parts of the chamber, accompanied by condensation or evaporation of fluid and the liberation or absorption of latent heat. This continues until the temperature variation is eliminated.

Just how small the temperature differences in a heat pipe are, even when it is not specially designed as an isothermal chamber, is illustrated by *fig. 9*. The diagram shows the temperature profile measured in the vapour duct of a stainless-steel heat pipe of 1 cm diameter. The working medium was sodium operating at a vapour pressure of about one atmosphere, and hence at a mean temperature of approximately 833 °C, the boiling point of sodium. The temperature drop along 50 cm of the pipe, from the beginning of the evaporator to the end of the condenser, is only a few degrees. (The sharp drop in temperature at the very end of the condenser is associated with the presence of a non-condensable gas in the heat pipe. We shall return to this subject in the section dealing with variable-conductance heat pipes.)

The dashed curve in fig. 9 is the temperature profile measured after the introduction of a quantity of mercury into the system. One effect of adding a second component to the working fluid is evidently the loss of the spatial isothermality. However, the behaviour of two-component systems is rather complicated, and will not be discussed here.

The most important industrial application of isothermal chambers is undoubtedly in ovens. A heat-pipe oven generally consists of two concentric cylinders between which a heat-pipe process can operate. Heating
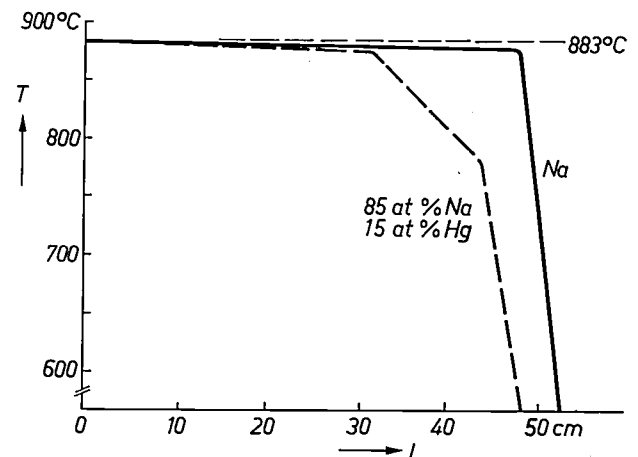
Fig. 9. Temperature distribution in a sodium-charged heat pipe. The temperature T in the vapour duct is plotted against the distance x between the point in question and the input end of the evaporator. It can be seen that the temperature drops only a few degrees over a length of 50 cm. The abrupt drop at the end is discussed in the final section of the article. The dashed line relates to an experiment in which the pipe contained a second condensable liquid, mercury, as well as sodium.

[*] Part I of this article appeared in Philips tech. Rev. **33**, 104-113, 1973 (No. 4).

[7] See for example: R. J. Meijer, Prospects of the Stirling engine for vehicular propulsion, Philips tech. Rev. **31**, 169-185, 1970.

[8] C. H. Dutcher Jr. and M. R. Burke, Electronics **43**, No. 4, 94, 16 Feb. 1970.

[9] C. A. Busse, Proc. 4th Intersociety Energy Conversion Engng. Conf., Washington D.C. 1969, p. 861.

is usually by electrical resistance, or is inductive, using radio-frequency coils. These may either surround the oven itself or deliver heat only to the evaporator section, which may be located at some distance from the oven. This permits a certain amount of flexibility, which is particularly important when the oven is to be enclosed in some other system. We shall now discuss some typical ovens in current use. Other types will be described in the final section of the article.

The black-body radiator

This type of oven was designed to produce black-body radiation for the calibration of pyrometers. Two versions were made, one operating at 450 °C and employing potassium as the working medium, and the other operating at about 750 °C and using sodium. Both are 13 cm long and have a diameter of 5.4 cm; they are heated by electrical resistance.

The isothermal insert

As the name indicates, this device was designed for insertion in a conventional oven to improve the uniformity of the temperature distribution (*fig. 10a*). It has no heating system of its own, but serves only to redistribute the heat from the surrounding oven. Ovens equipped with an isothermal insert are used at Philips in the measurement of physical properties of glasses. The working medium is again sodium, and the operating temperature lies between 700 °C and 850 °C.

The battery of ovens

Fig. 10*b* shows a number of ovens interconnected via a common evaporator region, so that they have a common temperature. Batteries of ovens are employed for softening copper wires in a continuous process, in which eight wires can be treated simultaneously. For this application potassium is the working medium, the operating temperature is about 450 °C and the length of the oven is about one metre.

Ovens with relatively high or low vapour pressure

Industrial processes often require conditions that are difficult or impossible to reconcile with the desirability of operating the oven at a temperature such that the vapour pressure is approximately equal to the pressure outside the heat pipe. In a sintering process, for example, the required temperature was 1150 °C. The natural choice of working medium would be lithium (see I, fig. 5 or Table III), but the corrosion problems associated with this metal are so great as to make the employment of sodium preferable. The vapour pressure of sodium at this temperature, however, is about 7 atmospheres, which introduces a new problem since above 1000 °C the creep strength of most metals falls rapidly.
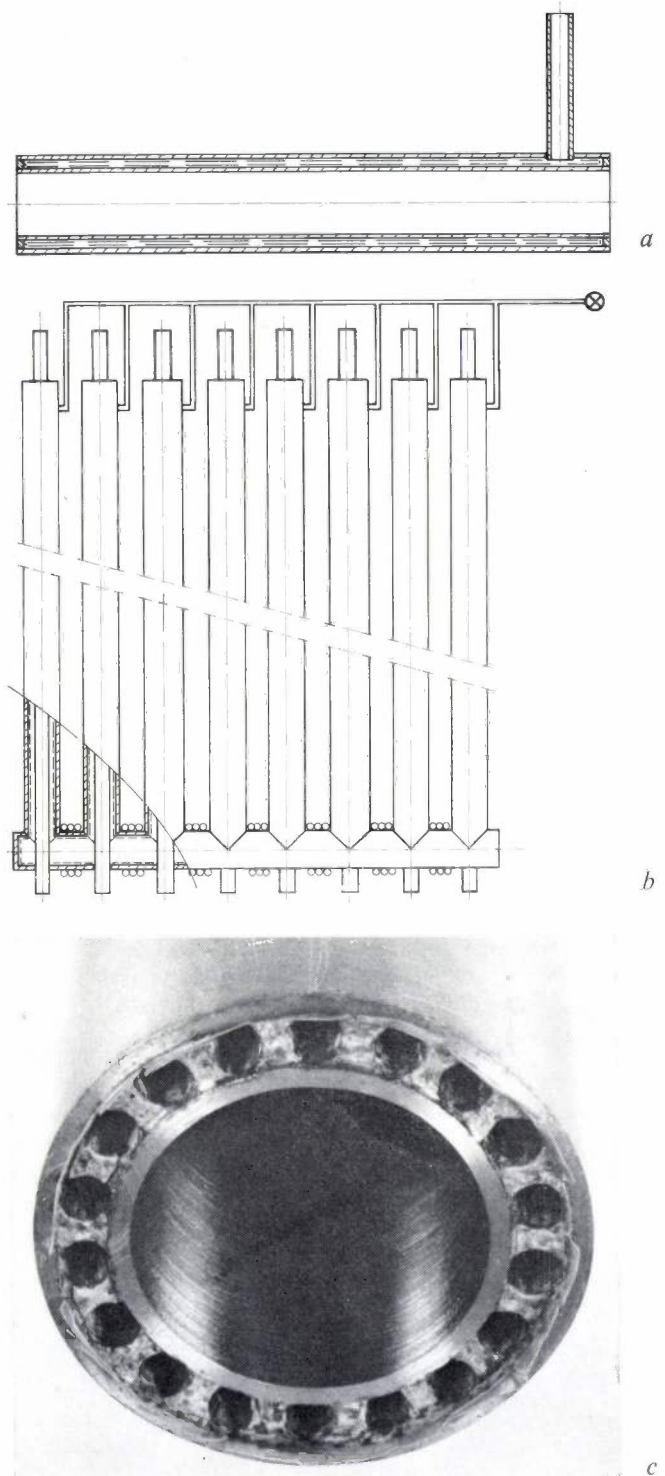
Fig. 10. Three examples of isothermal spaces operating in a specific temperature range. They consist in principle of two concentric cylinders, the outer one being the evaporator and the inner one the condenser; the heat-pipe process takes place in the space between them.
a) Isothermal insert. The whole pipe is placed in an oven to redistribute the heat more uniformly than is possible with the oven alone.
b) A battery of eight ovens interconnected by a common evaporator. These ovens are used for the simultaneous softening of eight copper wires.
c) A sodium-charged oven for 1150 °C. Since this temperature requires a sodium pressure of about 7 atmospheres, instead of one space acting as the heat pipe between inner and outer wall, as in (*a*), there are here 18 axial channels.

To minimize thermal creep in the oven, we employed an oven like the one shown in cross-section in fig. 10c. The oven consists of a thick-walled steel pipe (25/20 Cr-Ni steel) with channels drilled in the wall. Each channel contains a wick, and each works as a heat pipe in the *radial* direction. In this way all the channels together contribute to the heat transfer from the outside to the inside wall.

We adopted a similar solution for an oven in which the sodium vapour pressure in the space between inside and outside wall had to be much lower than the ambient pressure. This was required to be some tens of atmospheres for the sintering of optical fibres at a temperature of about 800 °C. The oven was constructed from 18/8 Cr-Ni steel and grooves were cut in the inside wall. These ensured that even if both walls were pressed together by the external pressure, the result would in fact be the cross-section shown in fig. 10c, and the heat-pipe function could still continue in the space left by the grooves.

Heat-flux transformers

The surface areas of the evaporation and condenser sections of a heat pipe need not be the same. Because of this a heat flux may be transformed from a higher to a lower value, or vice versa, since of course the total heat-transport rate has the same value at evaporator and condenser. This useful feature arises because the rates of evaporation and condensation are determined primarily by local conditions, the most important one being the difference ΔP between the pressure of the vapour and the saturation value. As we saw in I (page 109) this depends on the heat flux q and is given by:

$$q = \Delta P\, L\, (M/2\pi RT)^{\frac{1}{2}}.$$

Since this relation holds equally for evaporation and condensation processes, the ratio of the heat flux at the evaporator to the heat flux at the condenser — the 'transformation ratio' — is:

$$q_E/q_C = \Delta P_E/\Delta P_C,$$

where the subscripts E and C refer to the evaporator and condenser sections respectively.

Since the total heat transport is the same at evaporator and condenser, we have:

$$Q = A_E q_E = A_C q_C.$$

If the Clausius-Clapeyron equation is now used to relate pressure and temperature differences, we may write for the limit in which these differences are small:

$$q_E/q_C = \Delta P_E/\Delta P_C = \Delta T_E/\Delta T_C = A_C/A_E = R_{L-V}/R_{V-L}.$$

Since the flux-transformation ratio is equal to the *ratio* of two pressure differences, it may have a very large value even when the individual pressure differences are themselves small compared with the mean vapour pressure in the heat pipe.

This particular feature of heat-pipe systems is being exploited in several systems currently under development at Philips Research Laboratories. In this way it is possible to provide a Stirling engine with a heat-pipe flux transformer between the burner and the heater head. This will enable the heat-transfer surface on the burner side (the evaporator) to be much larger than that of the engine-heater pipes on which the working medium will condense.

Fig. 11 shows an experimental indirect heating system fitted to a Stirling engine. The heat pipe uses sodium as the working medium and operates at a mean temperature of 730 °C, transporting heat at an average rate of about 25 kW.

In the test arrangement in fig. 7 the heating is electrical, but in the engine itself the system is heated by atomized diesel fuel, resulting in a mean combustion-gas temperature of 1600 °C. Heat is transferred to the working medium of the heat pipe at an average rate of about 14 W/cm². The sodium vapour flows through
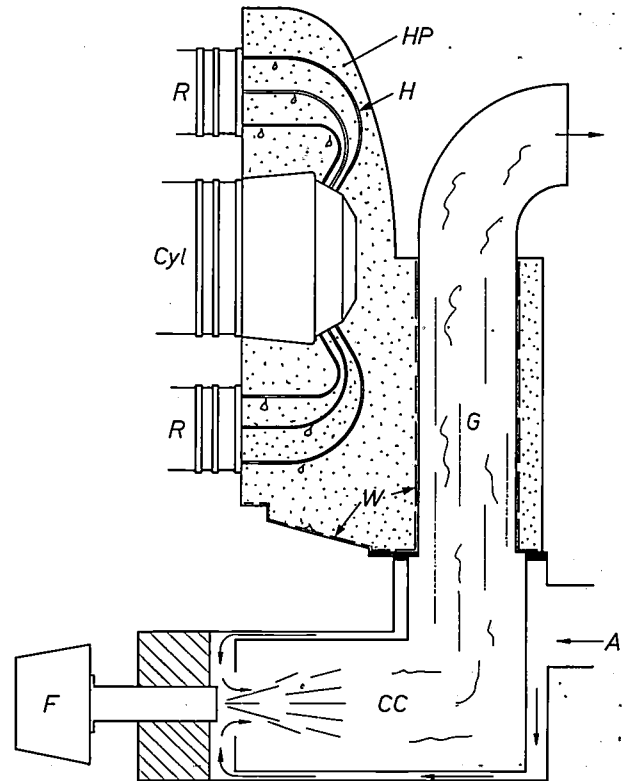
Fig. 11. Indirect heating system for a Stirling engine in which a sodium-charged heat pipe *HP* operates as a heat-flux transformer. Combustion gases deliver their heat to the heat pipe via a large surface (small heat flux), but this transfers the heat to the engine via a much smaller surface area (large heat flux). *Cyl* cylinder, *R* regenerator and *H* heater pipes of the engine. *W* gauze wick. *F* fuel atomizer. *CC* combustion chamber. *A* air inlet. *G* combustion gases.
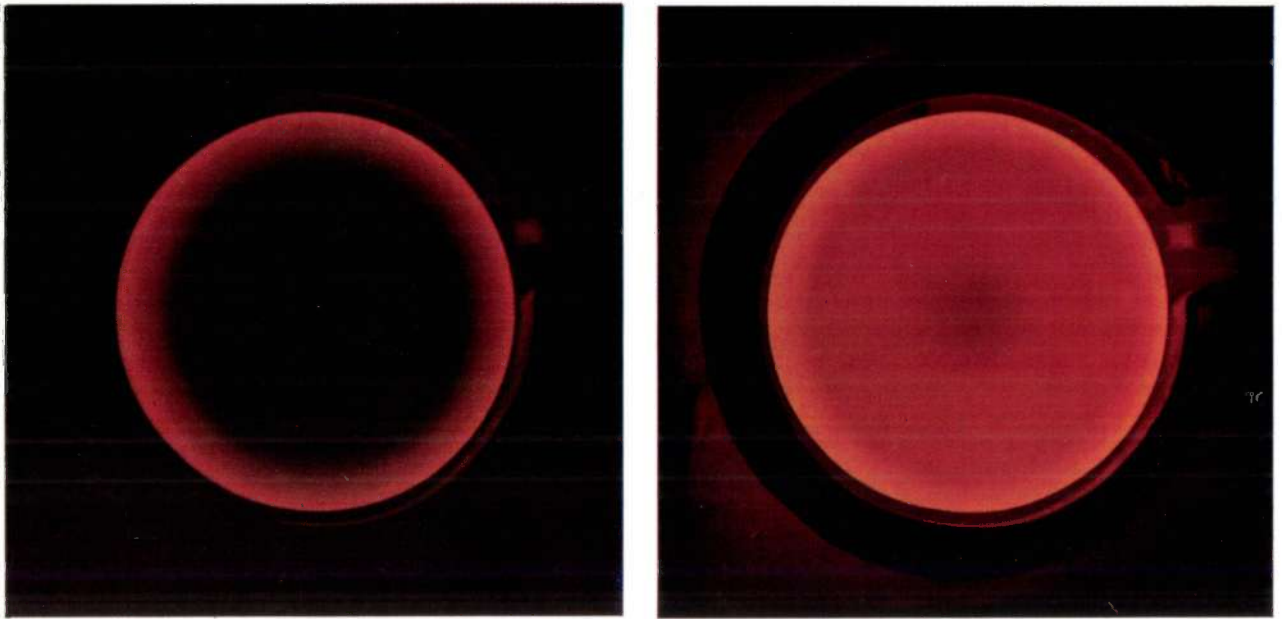
Fig. 12. Wickless heat pipe in the form of a horizontal hollow disc, containing sodium as the working medium. The disc is heated circumferentially by r.f. induction, and its edge therefore functions as the evaporator. When the disc is stationary (at the left) there can be no heat-pipe action, since the condensate does not return to the circumference. When the disc is rotated about a vertical axis (on the right), centrifugal force ensures the transport of the condensate and the disc then operates as a heat pipe.

the heat pipe at a rate that corresponds on average to an axial heat flux of 500 W/cm². At the engine head, sodium vapour condenses on the heater pipes of the engine, and here the heat flux is about 23 W/cm². The corresponding heat-flux transformation ratio is thus 1 : 1.6, and the calculated temperature drop between evaporator and condenser is 2.7 °C.

Another application of flux transformation that we have investigated concerns a wickless sodium-charged heat pipe; the pressure head necessary for the return of condensed liquid to the evaporator region is obtained by rotation of the heat pipe. This heat pipe, in the form of a hollow, rotating disc, was meant to simulate the anode of a 'Rotalix' X-ray tube. With rotating anodes it could be advantageous, for several reasons, if the heat generated at the small target area where the electron beam strikes it could be distributed uniformly over the whole anode disc and dissipated from there as radiation. If the disc itself consists of a heat-pipe system, and provided that the heat-pipe operating limits are not exceeded — an important factor in this instance — then the disc will function as a heat-flux transformer that converts a very high input flux into a low output flux.

Fig. 12 shows a test system in which the hollow disc is heated circumferentially by a small radio-frequency coil to simulate the localized nature of the heat input to a 'Rotalix' anode. Rotation of the disc produces heat-pipe action, with the outer edge acting as the evaporator.

Variable-conductance heat pipes

A class of heat pipes not yet discussed are those whose thermal conductance can be varied between very wide limits. The heat flow Q in these pipes depends to an exceptional extent on the mean temperature T_{Vm} of the vapour, or, in other words: dQ/dT_{Vm}, the temperature sensitivity, is very high. This feature, as we shall see later, is obtained by introducing a non-condensable inert gas into the heat pipe in addition to the working medium.

Variable-conductance heat pipes have an enormously wide range of potential applications, particularly in situations where the temperature is required to remain practically constant despite considerable fluctuations in the thermal loading. A typical example is the cooling of electronic components, whose temperature is required to be maintained within a specified range even though the power they dissipate during operation may vary a great deal. Before discussing examples of such heat pipes, however, we shall take a closer look at the temperature sensitivity itself.

Temperature sensitivity of a heat pipe

To see how the temperature sensitivity of a heat pipe is related to the design parameters, it will be useful to return to the description of a heat pipe as a group of thermal resistances connected in series, as shown in fig. 4 (Part I). For simplicity we shall lump together all the thermal resistances associated with the evaporator section and call them R_E. Similarly, R_C denotes the

sum of the resistances associated with the condenser section. As before, the temperature drop in the vapour duct is accounted for by a thermal resistance R_V. It will be evident that R_E is inversely proportional to the area of the evaporator, and R_C inversely proportional to the area of the surface where the condensation occurs.

We now consider the case of a heat pipe delivering an amount of heat Q to a heat sink at a fixed temperature (*fig. 13*). The temperature drops $\Delta T_{E,V,C}$ in the

of the vapour duct are small. The first can be achieved by having a large condenser area, or by ensuring that the liquid in the wick and also the envelope material have a high thermal conductivity.

We now consider a situation in which the condenser resistance R_C depends in one way or another on the amount of heat transported Q, and we shall see what the relationship should be to obtain the highest possible temperature sensitivity. It is easily shown that the thermal response is now given by:
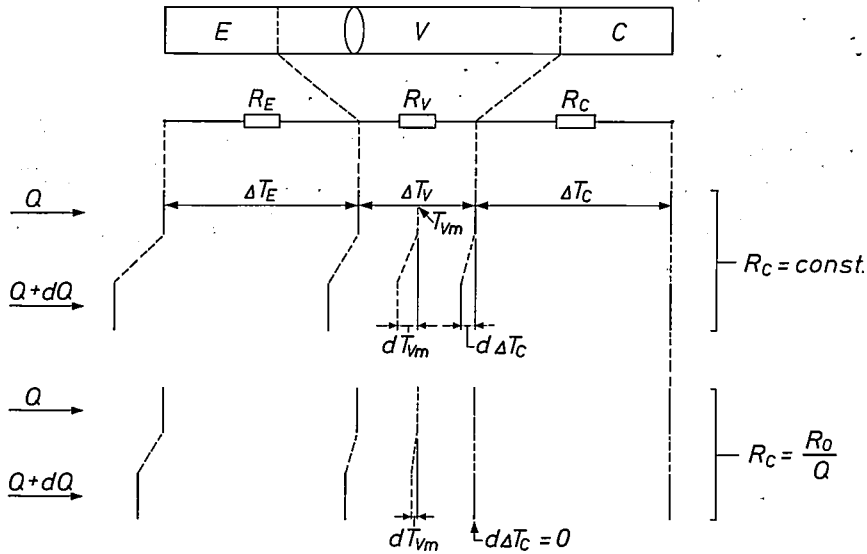
Fig. 13. The extent to which the mean temperature T_{Vm} in the vapour duct varies with the heat flow Q can be considerably reduced by designing the condenser section in such a way that its total thermal resistance R_C (see fig. 4: $R_C = R_{W(C)} + R_{Env(C)} + R_{Env-s}$) decreases as Q increases, and is at a minimum when R_C is inversely proportional to Q. R_E thermal resistance of evaporator. R_V thermal resistance of vapour duct. $\Delta T_{E,V,C}$ temperature drop across $R_{E,V,C}$.

heat pipe across each of these three resistances can then be written as three lumped thermal resistances:

$$\Delta T_E = QR_E; \quad \Delta T_V = QR_V; \quad \Delta T_C = QR_C.$$

Now suppose that the heat transported by the pipe is increased by an amount ΔQ while all the thermal resistances remain constant. The accompanying change in the mean temperature of the vapour (see fig. 13) will then be given by:

$$dT_{Vm} = d\Delta T_C + \tfrac{1}{2} d\Delta T_V,$$

and the 'thermal response' of the heat pipe by:

$$dT_{Vm}/dQ = d\Delta T_C/dQ + \tfrac{1}{2} d\Delta T_V/dQ = R_C + \tfrac{1}{2}R_V.$$

The temperature sensitivity is the reciprocal of this quantity, i.e.

$$dQ/dT_{Vm} = (R_C + \tfrac{1}{2}R_V)^{-1}.$$

A high temperature sensitivity is thus obtained when the thermal resistance of the condenser section and that

$$dT_{Vm}/dQ = d\Delta T_C/dQ + \tfrac{1}{2} d\Delta T_V/dQ =$$
$$= R_C(Q) + Q \, dR_C/dQ + \tfrac{1}{2}R_V.$$

The lowest value of the thermal response is $\tfrac{1}{2}R_V$ and is reached when $R_C(Q) = -Q \, dR_C/dQ$, or when $R_C = R_0/Q$, where R_0 is a constant. The maximum value of the temperature sensitivity is therefore:

$$dQ/dT_{Vm} = 2/R_V$$

and the temperature drop in the condenser is:

$$\Delta T_C = QR_C(Q) = R_0 = \text{constant}.$$

This is the situation shown at the bottom of fig. 13. The derivative $d\Delta T_C/dQ$ is zero here, which implies that the heat flux in the condenser is independent of the total heat transported by the pipe. In practice this situation will only be encountered if the effective part of the condenser surface area can be varied proportionately to the total heat transported.

We shall now discuss the method of obtaining a con-
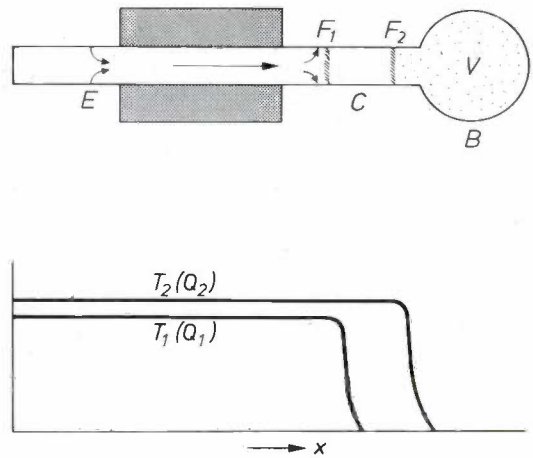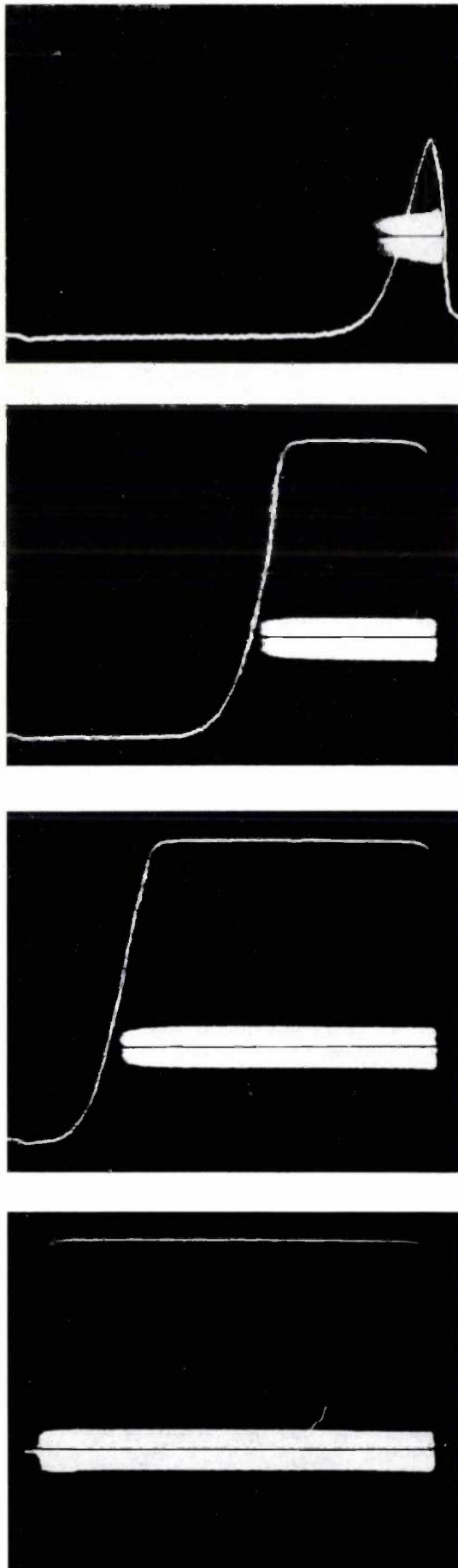
Fig. 14. Schematic representation (above) and temperature distribution (below; temperature T as a function of space coordinate x) for a heat pipe whose thermal response has been reduced by connecting the condenser to a buffer space B (volume V) and adding a non-condensable inert gas to the heat-pipe system. When the pipe is in operation, the gas is swept out of a large part of the space by the vapour flow, and a fairly sharply defined interface F is formed between the vapour and the gas. Only the part of the condenser to the left of F is effective. When the heat flow is increased from Q_1 to Q_2, the position of F shifts from F_1 to F_2, and the effective surface area of the condenser increases; the resistance R_C (see fig. 13) therefore decreases. When V is very large, R_C varies in inverse proportion to Q. At the interface there is a sharp drop in the temperature (the 'heat front'). In the equilibrium situation the vapour pressure is equal to the pressure of the buffer gas compressed in C and B. The buffer pressure thus controls the operating temperature.

denser surface with a variable effective area, and we shall see that the temperature sensitivity can be given any value between the maximum value just mentioned and the value found in a conventional heat pipe.

Gas-buffered heat pipes

The quantity R_C can be made to depend on Q by adding a non-condensable inert gas to the working medium in the heat-pipe system. This gas has the effect of buffering or blanketing off part of the condenser section of a heat pipe to a degree which can be made to depend upon the heat transported along the pipe. A gas-controlled heat pipe of this kind is shown diagrammatically in *fig. 14*, and we shall now explain its operation.

Fig. 15. Thermographs of a heat pipe containing water as the working fluid and a non-condensable gas, taken during warm-up. The evaporator is on the right. Power consumption was 130 W. These thermographs (taken at intervals of 20 seconds) were made with an instrument that carries out a horizontal line scan of the pipe and its environment [10]. Parts where the temperature is higher than 40 °C appear white, other parts black. The white curve is the temperature profile along the horizontal, which appears black in a picture of the pipe. It can be seen that the vapour/gas interface is fairly sharply defined and that it moves towards the condenser as the heat pipe warms up. In the part to the right of the interface, where heat-pipe action takes place, the temperature along the axis of the pipe varies by only about 0.1 °C.
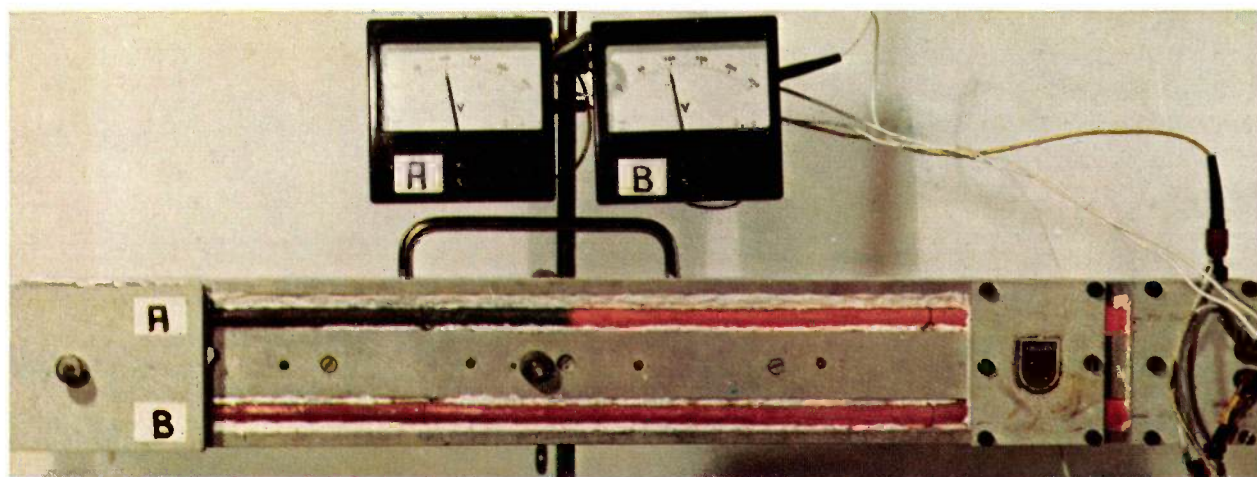
Fig. 16. Sodium-charged heat pipes with identical electric heaters and the same input power. The evaporators are on the extreme right. In addition to sodium the upper pipe contains argon (approx. 200 torr); the lower one is a conventional heat pipe. At the front the insulation has been removed over some distance so that this part acts as the condenser. The lower pipe dissipates the power input across the whole visible length, the other one only to the right of the heat front, and therefore at a higher temperature.

As can be seen in the figure, the condenser section C of the heat pipe is connected to a buffer volume B. When the pipe is not in operation, gas and vapour are of course homogeneously mixed, and the total pressure in the vapour duct is equal to the sum of the two partial pressures. When the pipe is in operation, however, the situation is quite different. Owing to the flow of vapour from evaporator to condenser, the non-condensable gas is swept along in the vapour stream, and compressed into the buffer space and in the adjacent part of the condenser section. The interface between the part of the heat pipe filled with vapour and the part filled with gas is fairly sharply defined, and in this situation the vapour pressure is therefore approximately equal to the pressure of the compressed buffer gas. Since the buffer gas cannot condense it cannot be recycled and hence plays no part in the heat-pipe process, which takes place exclusively in the vapour-filled part. The effect of increasing the vapour flow is to compress the buffer gas even more, causing a larger surface area to become available for condensation, so that in fact R_C decreases as Q increases. If the buffer volume is made very large, then as we shall see in more detail below R_C is inversely proportional to Q and the temperature sensitivity reaches its maximum value.

The marked separation between buffer gas and vapour in a gas-controlled heat pipe is demonstrated in *fig. 15*. These photographs relate to a heat pipe filled with water and nitrogen. *Fig. 16* shows that the situation at higher temperatures is no different. This compares two sodium heat pipes, both at equilibrium, but with one gas-controlled and the other operating in the normal mode.

Both heat pipes transfer the same total amount of

heat from evaporator to condenser. The conventional heat pipe, however, can give up its heat over the whole of the condenser section, and thus operates at a lower mean temperature than the gas-controlled pipe, which has part of its condenser section blanketed off by argon. In the gas-controlled pipe the temperature drop at the interface is marked by the abrupt fall in the radiation emitted from the heat-pipe wall.

An interface as sharply defined as this, in a plane perpendicular to the axis of the heat pipe, is found only in pipes of relatively small cross-sectional area and with large vapour flow rates. If the pipe is upright or sloping however, with the condenser uppermost, and if the gas is heavier than the vapour, then fairly considerable mixing of gas and vapour is likely to occur if the amount of heat transported is small. This invariably means a blurring of the interface and also a loss of isothermality in the evaporator and transport sections.

A somewhat less dramatic situation is demonstrated in *fig. 17*, which shows a sloping stainless-steel heat pipe of 5 cm diameter, containing water as the working fluid and helium as the buffer gas, where the interface, though inclined, is nevertheless fairly sharply defined.

An important question for the designer is of course how large the volume V of the buffer space connected with the condenser section (fig. 14) should be if the mean working temperature is to remain between the limits T_1 and T_2 when the heat flow is varied between the limits Q_1 and Q_2.

In the usual theoretical treatment of the gas-con-

[10] This equipment was developed in the Laboratoires d'Electronique et de Physique Appliquée (LEP) at Limeil-Brévannes, France, where we also took the photographs shown here. A description of the equipment is given in: M. Jatteau, Philips tech. Rev. **30**, 278, 1969.

trolled heat pipe, the heat front is assumed to lie in a plane perpendicular to the longitudinal axis of the pipe, and to be independent of the orientation of the pipe. Moreover the interface is assumed to be sharp. With these assumptions W. Bienert [11] has succeeded in expressing the operation of a gas-controlled heat pipe by a differential equation.

Using a simpler treatment an expression for V can be derived that proves to be sufficiently accurate for design calculations. This is:

known area of the condenser section required to dissipate the transported heat. For a cylindrical pipe of uniform cross-section (radius a, condenser length l) the following equations apply:

$$Q_1 = 2\pi a q_1 l, \qquad Q_2 = 2\pi a q_2 (l + \Delta l),$$

$$\Delta l = \frac{l}{2\pi a}\left[\frac{Q_2}{q_2} - \frac{Q_1}{q_1}\right],$$

$$\Delta V = \pi a^2 \Delta l = \frac{a}{2}\left[\frac{Q_2}{q_2} - \frac{Q_1}{q_1}\right].$$

Substitution in our equation for $V/\Delta V$ then yields the expression given above for V.
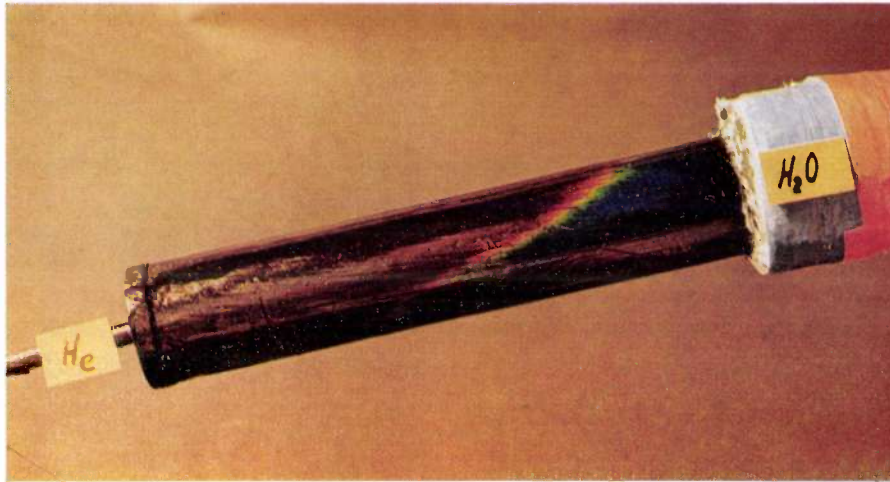
Fig. 17. Sloping heat pipe containing water and helium with evaporator above the condenser. The interface between vapour and gas has been made visible by covering the outside with a liquid-crystal layer that gives selective reflection as a function of temperature. Between about 60 °C and 65 °C the wavelength of the reflected light traverses the whole visible spectrum. The strip of condenser surface where the temperature passes through this interval therefore appears in the colours of the spectrum. It can be seen that the interface between vapour and working fluid, although askew, is still fairly sharply defined.

$$V = \frac{a}{2}\left[\frac{Q_2}{q_2} - \frac{Q_1}{q_1}\right]\left[\frac{P_2 t_1}{P_1 t_2} - 1\right]^{-1}.$$

Here $P_{1,2}$ are the saturated vapour pressures corresponding to the temperatures $T_{1,2}$, and $q_{1,2}$ is the heat flux through the surface area of the condenser at these temperatures; $t_{1,2}$ represents the mean temperature of the gas in both cases and a is the radius of the pipe.

This expression is arrived at as follows. In the equilibrium state, i.e. with no movement of the interface, the pressure of the gas is equal to the vapour pressure of the working fluid. Assuming that the buffer gas behaves like an ideal gas, then we have the following relations:

$$P_1(V + \Delta V) = R t_1 \quad \text{for } Q_1,$$

$$P_2 V = R t_2 \quad \text{for } Q_2.$$

Here ΔV is the volume swept out by the interface in accommodating the change of heat input from Q_1 to Q_2. Combining the above equations gives:

$$\frac{V}{\Delta V} = \left[\frac{P_2 t_1}{P_1 t_2} - 1\right]^{-1}.$$

The quantity ΔV may be independently determined from the

The above result was applied in the design of a heat-pipe oven that was not only required to be isothermal along its length (nearly two metres), but also had to be capable of accommodating heat-transport fluctuations of up to 10% (in this case one kilowatt) with only a 5 °C temperature variation about a mean vapour temperature of 825 °C.

The oven was designed so that the excess heat input could be dissipated by radiation. The vapour/gas interface was located in an uninsulated section of the pipe, and advanced or receded as the oven heat load varied. The oven design specification required a buffer gas pressure of 5.7×10^4 Pa and a buffer volume of about 1 litre. To be on the safe side, a volume of several litres was used in practice.

The equation for the required buffer volume V is considerably simplified if the temperature of the buffer gas can be considered constant, that is if $t_1 = t_2$. If at the same time $T_1 = T_2$, then $P_1/P_2 = 1$, and the required buffer volume is infinite. For practical purposes this means that it must be large compared with the pipe itself. The heat pipe is then operating in the

constant-pressure mode, which corresponds to its maximum temperature sensitivity.

As mentioned earlier, this maximum temperature sensitivity occurs when the thermal resistance of the condenser section is inversely proportional to the transported heat Q, or in other words when the effective area of the condenser section is proportional to Q. It can

isothermal plane surface at a known temperature. This temperature also had to be reached quickly and then held constant for long periods of time in spite of possible fluctuations in the power supply to the hot surface.

The main part of the equipment [12] (*fig. 18*) is an isothermal chamber *Ch*, which forms part of a gas-
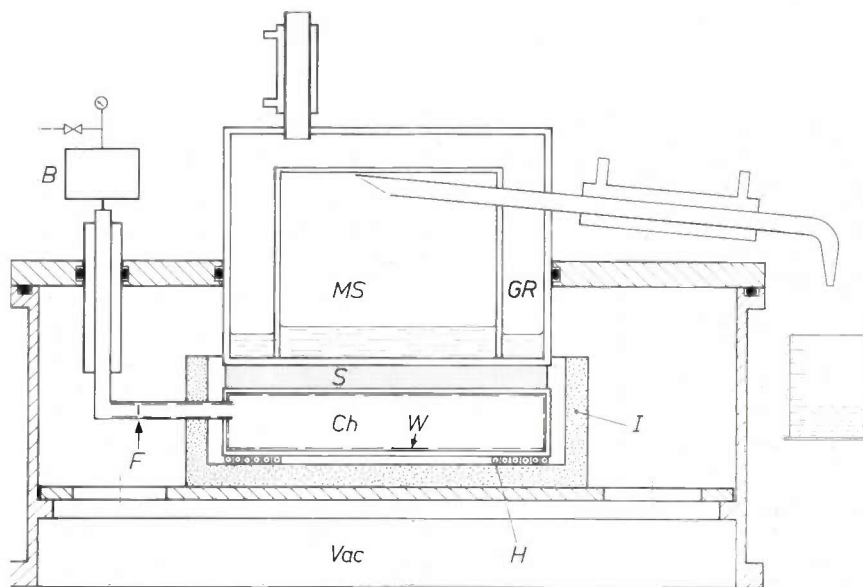
Fig. 18. Equipment for measuring the thermal conductivity of insulating material. Underneath the sample *S* is an isothermal surface of virtually constant temperature. *MS* measuring space. *GR* thermal guard ring. *Ch* isothermal chamber filled with sodium vapour. *W* wick (steel gauze). *H* heater. *I* insulation. *B* buffer gas volume. *F* position of heat front. *Vac* evacuated space. When the equipment is operating the liquid in *MS* boils, giving a constant and even temperature at the upper side of the sample. The heat flow through *S* is derived from the quantity of condensed liquid flowing out of the cooler in unit time.

readily be shown that this is in fact the case for a gas-controlled heat pipe operating in the constant-pressure mode.

For heat-pipe applications in which non-flammable working fluids are employed, it is clear that air is a very convenient buffer gas. Moreover, when a working pressure of about 1 atmosphere is required, the environment can provide an infinite buffer volume. The chemist's reflux condenser is in fact an example of this. In other instances, particularly where the working fluid is a liquid metal, argon is generally used as the buffer gas, and for constant-pressure operation a sufficiently large buffer volume is attached to the system. We have used an arrangement of this type employing liquid sodium and argon in an equipment for measuring the thermal conductivity of insulating materials, which we shall now discuss.

An equipment for measuring thermal conductivity

To measure the extremely low thermal conductivities of multiple-foil insulations we needed an equipment in which heat could be applied to the sample through an

buffered heat-pipe system. The chamber consists of a cylindrical nickel-chromium steel vessel, whose inside walls are clad with a wick formed from several layers of steel gauze spot-welded into position. A horizontal pipe, also containing a wick, protrudes from the vessel and is connected via a vertical arm to the gas buffer space *B*.

The isothermal chamber is put into operation in the following way. The pressure of the gas in the buffer space *B* is set at a value equal to the vapour pressure of the working fluid at the desired chamber temperature. The chamber is then heated electrically until the vapour pressure of the sodium equals that of the buffer gas, and evaporation occurs at a rate sufficient to displace the gas from the chamber and drive it into the horizontal pipe. Here a clear interface, or heat front, is developed between the argon and sodium vapour.

[11] W. Bienert, Proc. 4th Intersociety Energy Conversion Engng. Conf., Washington D.C. 1969, p. 1033.

[12] D. B. Green and H. H. M. van der Aa, 11th Int. Conf. on Thermal Conductivity (extended abstracts), Albuquerque, N.M., USA, 1971, p. 31. A somewhat similar apparatus for measuring medium to very high thermal conductivities was described by J. Schröder, Philips tech. Rev. 21, 357, 1959/60.

The sodium condenses continuously and is recycled to the evaporator area via the wick.

The position of the heat front in the horizontal arm is determined primarily by the area required to radiate the excess power input. As we noted earlier, the liquid behind the heat front in the wick is everywhere in equilibrium with its vapour, and since there are only very small pressure differences the chamber wall acquires almost the same temperature everywhere. The volume of the buffer vessel B is made sufficiently large to ensure that a change in the heat supply to the chamber Ch causes very little change of pressure, and therefore does not give rise to any appreciable change of temperature. The temperature of the wall of Ch in contact with the sample drifts less than 1 °C per hour.

a

b

c
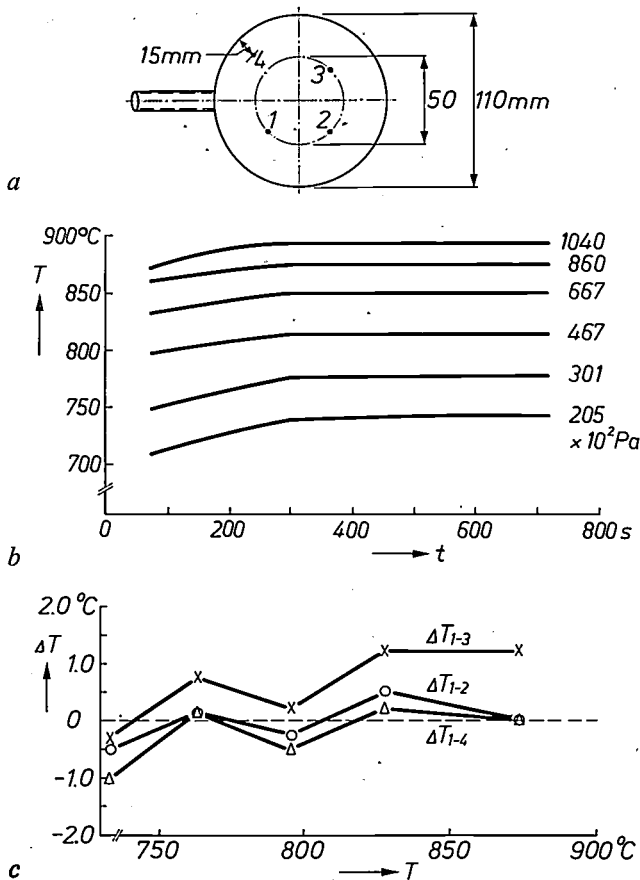
Fig. 19. Thermal characteristics of the isothermal chamber of the equipment shown in fig. 18, recorded by means of four thermocouples located in the upper plate.
a) Position of the four thermocouples, indicated by the numbers 1-4.
b) Temperature variation at the location 1 during warm-up measured at six different initial gas pressures.
c) Temperature differences between the location 1 and the three others, plotted against the mean chamber temperature at which the measurements were made.

Fig. 19 shows some of the results obtained. The measurements were made with the aid of four thermocouples located at the positions marked 1-4 in fig. 19a. The warm-up characteristics of the chamber (fig. 19b) were monitored by thermocouple 1 for various values of the buffer-gas pressure. It can be seen that the top-plate temperature rises at first until after about 5 minutes the vapour pressure is equal to that of the buffer gas; no further temperature change is then recorded.

After this state has been reached, the temperature gradients in the plate are generally less than 0.5 °C/cm, as the curves in fig. 19c show. This figure shows the temperature differences existing between the various thermocouples, for various temperatures between 730 °C and 875 °C. The maxima of these differences barely exceed one degree.

The examples discussed above show that the high temperature sensitivity that can be obtained for a gas-buffered heat pipe can lead to improved solutions to many problems both in scientific instrumentation and in manufacturing processes.

In this article we have tried to give an outline of the principles underlying the operation of heat pipes, and we have discussed some applications in which the emphasis was placed on those principles and features that make the heat pipe so useful. When the four characteristic features of heat pipes considered here are taken together, it is clear that devices based on them pose a formidable challenge to many of the existing methods of heat transportation, heat distribution and temperature control.

Summary. This part of the article discusses examples of heat-pipe systems that function as transporters of heat (in Stirling-engine experiments and in a siphon for the transfer of molten salts), as an isothermal space (black-body radiator, ovens, isothermal inserts for ovens, operating at both high and low vapour pressures) and as heat-flux transformers (in a Stirling-engine heating system using combustion gases and in a rotating anode for X-ray tubes). When a non-condensable inert gas is added to the working medium in a heat-pipe system, the thermal resistance of the condenser can be made to depend on the quantity of heat Q transported along the pipe, thus considerably reducing the extent to which the mean temperature of the vapour varies with Q (temperature sensitivity). This reduction depends on the volume of a buffer space connected with the condenser section. In a gas-buffered heat pipe the vapour compresses the gas into the buffer space and the adjoining part of the condenser section, so that this part of the condenser section takes no part in the heat-pipe process. The position of the vapour/gas interface moves as the heat flow varies. The article discusses an application of this principle in an equipment for measuring the thermal conductivity of insulating materials.

ERRATUM. In Part I of the above article, Philips tech. Rev. 33, 1973 (No. 4), the upper equation in the left-hand column of p. 110 should read: $\Delta P_v = 4\eta vql/(L\varrho v r_v^2)$.

PHILIPS TECHNICAL REVIEW

A simple model for alloys

A. R. Miedema

I. Rules for the alloying behaviour of transition metals

II. The influence of ionicity on the stability and other physical properties of alloys [*]

Reliable rules for the alloying behaviour of metals have long been sought. There is the qualitative rule that states that the greater the difference in the electronegativity of two metals, the greater the heat of formation — and hence the stability. Then there is the Hume-Rothery rule, which states that two metals that differ by more than 15% in their atomic radius will not form solid solutions. This rule can only be used reliably (90% success) to predict poor solubility; it cannot predict good solubility. The author has proposed a simple atomic model, which is empirical like the other two rules, but nevertheless has a clear physical basis and predicts the alloying behaviour of transition metals accurately in 98% of cases. The model is very suitable for graphical presentation of the data and is therefore easy to use in practice.

I. Rules for the alloying behaviour of transition metals

Introduction

The fusion of two metals can lead to widely different results. In some instances fusion is in fact not possible because the metals in the molten state refuse to mix at any reasonable temperature, rather like oil and water. Examples of such an immiscible system of two metals are iron and lead and iron and bismuth. It may also happen that the two metals are miscible as liquids but not in the solid phase. When silver and copper are melted together, the result after cooling is a eutectic mixture of silver (with a few per cent of copper dissolved in it) and copper (with a small proportion of dissolved silver). A third possibility is that the two metals may readily form intermetallic compounds. The alloying of nickel with titanium produces compounds with a crystal structure that differs from that of both Ni and Ti. Just which compound is obtained depends on the atomic ratio in which Ni and Ti were alloyed. During the formation of an intermetallic compound a great deal of heat may be released, and indeed the alloying of two metals — Hf and Pt — has caused an explosion [1].

Are there any rules that could be used to predict the alloying behaviour or reactivity of two metals with some degree of certainty? Some rules have indeed been proposed in the past. For example, it has been stated in the literature [2] [3] that two metals more readily form intermetallic compounds the more they differ in their electronegativity.

Electronegativity is a frequently used term in inorganic chemistry, and was introduced by Linus Pauling. In principle, electronegativity is a measure of

---

[*] Part II will appear in the next issue.
[1] L. Brewer, Acta metall. **15**, 553, 1967.
[2] W. Hume-Rothery, in: Phase stability in metals and alloys (ed. P. S. Rudman, J. Stringer and R. I. Jaffee), McGraw-Hill, New York 1967, p. 3.
[3] O. Kubaschewski, E. Ll. Evans and C. B. Alcock, Metallurgical thermochemistry, 4th edition, Pergamon, Oxford 1967.

Dr A. R. Miedema is with Philips Research Laboratories, Eindhoven.

the affinity of a particular chemical element for electrons. Strongly electronegative elements tend to form negative ions (oxygen and fluorine), and the less electronegative elements tend to form positive ions in chemical compounds. Thus the electronegativity of metallic elements can be measured in terms of their reactivity to oxygen, for example. The weakly electronegative metals form stable, ionic oxides; the reactions are strongly exothermic, i.e. the heat of formation is strongly negative. The more electronegative (noble) metals, on the other hand, differ less from oxygen in their electronegativity: the oxides of these metals are not very stable.

When two metals of distinctly different electronegativity are melted together, the result may be expected to be a somewhat 'ionic' alloy. The alloy AB, formed from the metals A and B, can be described as $A^{+\Delta Z} B^{-\Delta Z}$, where ΔZ indicates the charge transfer per atom and can have an arbitrary value (which does not necessarily have to be an integer). The greater the difference in electronegativity between the metals A and B, the larger is the value of ΔZ, and the greater therefore will be the heat of formation and the stability of the alloy.

This is illustrated in *fig. 1* by a comparison of the phase diagram of Ca and Ag with that of Ca and Au. The two phase diagrams closely resemble one another and show a number of intermetallic compounds; the melting points of the Ca-Au compounds, however, are always substantially higher than those of the corresponding Ca-Ag compounds. This agrees with the difference in electronegativity: Ca is the least electronegative element of the three, and Au the most electronegative, so that the Ca-Au system yields more stable compounds with higher melting temperatures. Another example is given in *fig. 2*. In the Au-Pb system two compounds occur that do not exist in the Ag-Pb system because of the small difference in electronegativity

between Ag and Pb. A similar situation is encountered when the phase diagrams of Au and Ag with Ga, Sn, Bi, Ta, V and Cu are compared.

Although metallurgists are generally agreed that the difference in electronegativity is an important factor in determining whether or not two metals will alloy, it is not easy to derive any quantitative rule from this, firstly because electronegativity is certainly not the only factor involved, and secondly because it is not clear from which physico-chemical properties the electronegativity can best be deduced. There are various scales [4-7] for the electronegativity of the elements, in which more than one value is occasionally given for one element, depending on the valency. (What then is the valency of a metal?) The various scales agree in general terms, but there are some significant differences in detail. Electronegativity remains in any case a somewhat vague concept, which is never encountered in more physical descriptions of metals.
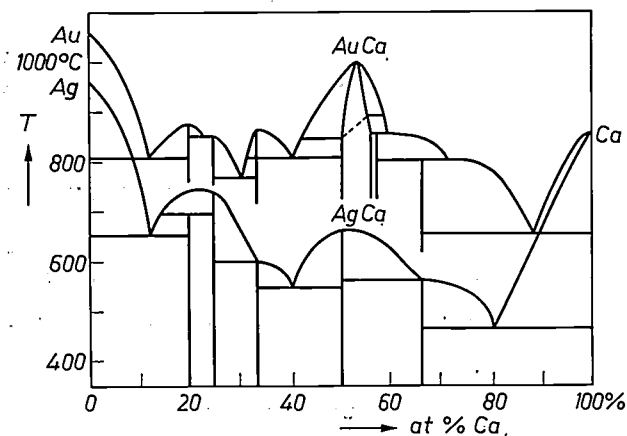
Fig. 1. Comparison of the phase diagrams of calcium and silver with that of calcium and gold. The greater difference in electronegativity between Ca and Au leads to relatively higher melting points (after W. Hume-Rothery [2]).
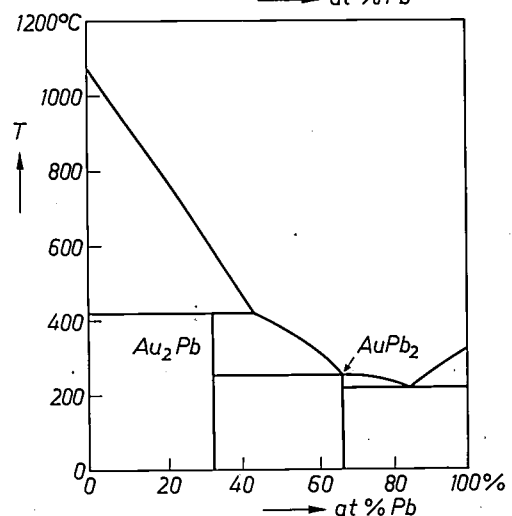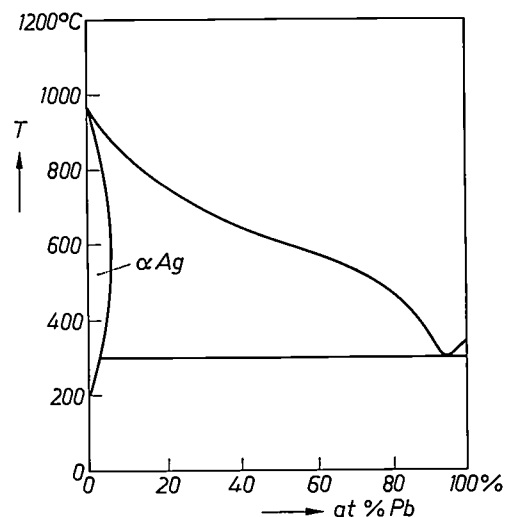
Fig. 2. Comparison of the phase diagrams Ag-Pb and Au-Pb. Because of the minimal difference in electronegativity between Ag and Pb, two compounds are not found in the system Ag-Pb that do occur in the system Au-Pb.

Fig. 3. Darken-Gurry diagram, the electronegativity X plotted against the atomic radius r_{12}, for iron (after J. J. Waber *et al.*[8]). The rule here is that the elements that dissolve well in iron lie inside the ellipse around Fe with the axes 0.20 Å atomic-radius difference and 0.4 difference in electronegativity. The other metals wil not readily dissolve (R.E. represents the elements of the rare-earth group.)

The scale used for the electronegativity X is that given by E. Teatum *et al.*[6]. It is shown later on in this article that the best electronegativity scale for metals differs slightly from this.
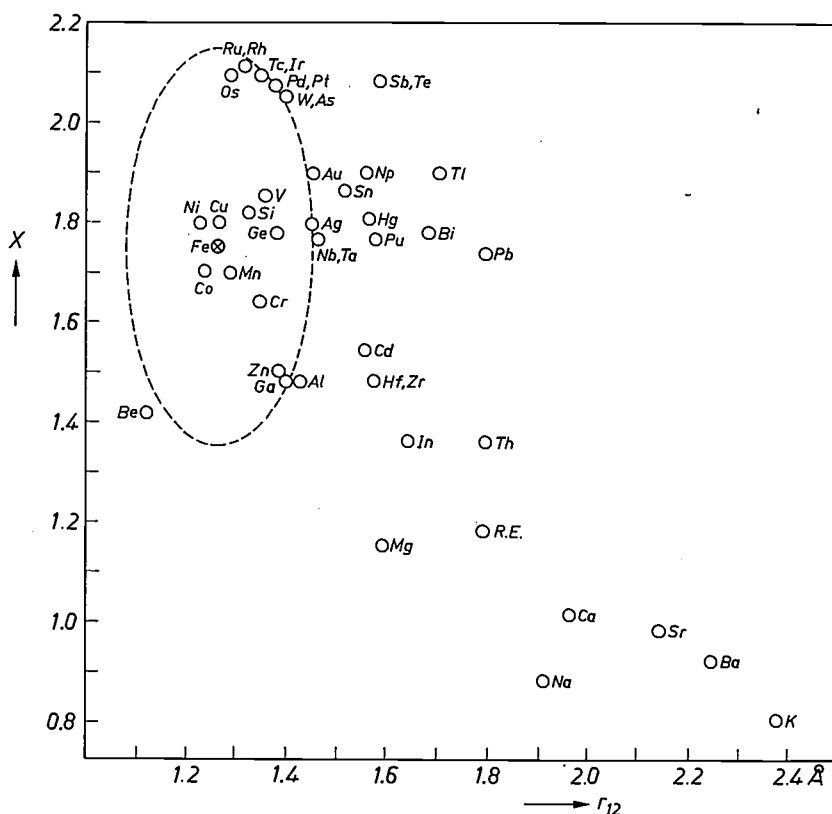
Another rule that says something about the alloying of metals is known as the Hume-Rothery rule. This states that two metals that differ by more than 15% in their atomic radius will not form solid solutions. Atoms differing considerably in size will not readily form a mixed crystal (e.g. a phase with the face-centred cubic, body-centred cubic or hexagonal close-packed structure [*] in which the atoms are statistically distributed among the lattice sites), but can of course be accommodated in an intermetallic compound, in which the different metal atoms occupy crystal sites that are not equivalent. The 15% rule explains, for example, why Fe and Bi do not mix (the atomic volumes differ by a factor of 3) or why the two f.c.c. metals Cu and Ag do not easily dissolve in one another (Cu is much smaller than Ag). An extensive study of the applicability of the 15% rule, made from all the binary phase diagrams known in 1962 was carried out by J.T.Waber *et al.* [8]. They defined the solubility of metal A in metal B as good if it was greater than a fixed value which they took to be 5 at %, and poor if it was less than this value. The analysis given by Waber *et al.* showed that the rule was quite effective when it was a matter of predicting the *poor* solubility of two metals (out of 619 cases of poor solubility 559 were correctly predicted, i.e. 90%), but that it was useless for predicting *good* solubility (out of 804 cases 403 were correctly predicted, i.e. 50%). Waber *et al.* also showed that the prediction of solid solubilities can be improved by combining the electronegativity factor with the difference in atomic size (the size factor).

They used a rule that was formulated in principle by L. S. Darken and R. W. Gurry. The basic postulate is that two metals do not dissolve in one another if the size factor is too large, or if the difference in electronegativity is too large. If the electronegativity difference is too large there is a chance of intermetallic compounds forming, which will compete in stability with the solid solution, and thus reduce the solubility. This leads to a simple geometrical method (*fig. 3*) of predicting by inspection whether or not the various metals will dissolve in a particular matrix metal (in fig. 3 this is Fe). Two parameters — the electronegativity and the atomic radius — have been applied for each metal, so that each element corresponds to a point in fig. 3. The rule postulated by Waber *et al.* now becomes: metals that are readily soluble in Fe lie by definition within the ellipse that has Fe as the centre-point, whose horizontal axis corresponds to a 15% difference in atomic radius and whose vertical axis corresponds to a difference of 0.4 units in electronegativity. The metals that do not dissolve well in Fe lie outside the ellipse. This rule has

[4] L. Pauling, The nature of the chemical bond, 3rd edition, Cornell Univ. Press, Ithaca 1960.

[5] W. Gordy and W. J. O. Thomas, J. chem. Phys. **24**, 439, 1956.

[6] E. Teatum, K. A. Gschneidner Jr. and J. T. Waber, U.S. Atomic Energy Comm. Rept. LA-2345 (Los Alamos Sci. Lab.), 1960.

[7] R. T. Sanderson, Inorganic chemistry, Reinhold, New York 1967, p. 78.

[8] J. T. Waber, K. Gschneidner Jr., A. C. Larson and M. Y. Prince, Trans. Metall. Soc. AIME **227**, 717, 1963.

[*] The abbreviations f.c.c., b.c.c. and h.c.p. will be used in the rest of the article.

given a success rate of 77 % in predicting whether metals dissolve in one another or not.

The difficulties inherent in the Darken and Gurry method are evident from fig. 3. What is the best value for the electronegativity of a metal? What is the best parameter for expressing the size of atoms? The atomic radius as used in fig. 3 — taken as the value that would be found in a crystal with f.c.c. or h.c.p. structure — is a relatively arbitrary measure. A more fundamental objection is that a prediction with a probability of 77 % is still far from satisfactory, and in practical cases does not offer adequate certainty. What is more, failure to dissolve is not the same as failure to alloy. No indication is given as to how it could be possible for two metals that do not differ in atomic size, but do differ in electronegativity, not to be capable of forming an alloy (either solid solutions or compounds), as has been observed for Mg and Zr.

In this article a simple atomic model will be introduced [9] that leads to rules for predicting the alloying of two metals with a certainty of about 98 %. Although the model is essentially empirical, it does have a physical background, and the two parameters to be introduced have a distinct physical meaning. The consequences go further than the formulation of rules for the prediction of alloying behaviour. This will become evident when we later come to consider the work functions of pure metals and quantities such as the magnetic properties of the alloys of transition metals.

The model

· The essential starting point of the model used here is that the concept of individual atoms remains of significance for alloys. This differs from the usual theoretical treatment of metals, in which the starting point is a lattice of positive ions surrounded by a 'sea' of electrons. For pure metals it is customary to assign to each lattice point an atomic cell whose shape is such that all the atomic (Wigner-Seitz) cells together fill the entire space. In the case of an element with a simple cubic crystal structure the Wigner-Seitz cells will be cubes; in the two-dimensional example given in *fig. 4a* the atoms of metals A and B are represented as squares. It has been shown theoretically for pure metals that the shape of the atomic cell has little effect on the bond energy of a metal as long as the volume of the cell does not change. This means that in fact the crystal structure assumed by a metallic element is not very relevant to the bond energy (by which is meant the latent heat of vaporization at absolute zero). We see therefore that various metals can generally assume more than one crystal structure as a function for example of temperature or pressure (allotropy).
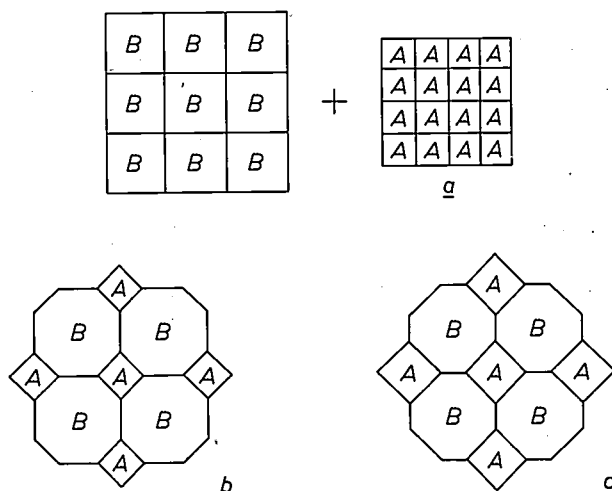
Fig. 4. Model for metals (*a*) and alloys (*b*). The alloy AB is considered to be built up from atomic cells that resemble those of the pure metals A and B. Equalization of the chemical potential for electrons gives rise to electron transfer, involving some change of volume (*c*). Transfer takes place from B to A.

As can be seen from fig. 4, it is assumed in the model that the atomic cells of A and B in the alloy closely resemble, as a first-order approximation, the cells of the two metals as separate elements. Although the shape has to be altered slightly, since otherwise the cells for the alloy AB (fig. 4*b*) would not take up the entire space, there is no change in the two atomic volumes, so that the energy effect need not be taken into account.

We shall now deal with two effects that correct the 'construction' of an alloy as just described. In general the chemical potential for electrons, Φ^*, will not be equal in the atomic cells A and B. This cannot remain so. Some electron transfer will take place from the cells with the relatively high chemical potential (B) to the sites with the lower, more favourable chemical potential (A atoms), in which process the cells B will decrease slightly in volume while the cells A will become bigger (fig. 4*c*). This transfer of electrons will always reduce the energy of the system. A contribution would therefore be expected to the heat of formation — i.e. at $T = 0$ the formation enthalpy ΔH of the alloy — which depends on the difference $\Delta\Phi^*$ in chemical potential for the two metals and which will always be negative. In a series expansion in $\Delta\Phi^*$ the first term will therefore be negative and proportional to $(\Delta\Phi^*)^2$.

A second correction to be made is connected with the fact that there must be no discontinuities in the density of the electrons (i.e. the number of electrons per unit volume). When the atomic cells of the pure metals A and B are combined to form an alloy, the electron density n_{WS} at the boundary of a cell A will not generally be equal to that at the boundary of a cell B. This discontinuity should be eliminated. It is possible to do

this by compressing the cell with the lower density and letting the cell with the higher density increase in volume. Since the original atomic volume of the metals A and B corresponded to an energy minimum, any change makes a positive contribution to the heat of formation of the alloy. One may expect a contribution to ΔH that will depend on the difference in the density of electrons at the boundary of the two atomic cells, Δn_{WS}, and this will always be positive. In a series expansion in powers of Δn_{WS} the first term is therefore positive and proportional to $(\Delta n_{WS})^2$.

The model, as outlined in principle in fig. 4, results in the following prediction for the heat of formation of alloys:

$$\Delta H = f(c) \, [-Pe(\Delta \Phi^*)^2 + Q(\Delta n_{WS})^2], \qquad (1)$$

where $f(c)$ is a still unspecified function of the concentration of the metals, e is the electronic charge, while P and Q are assumed to be constants.

If P and Q are in fact constants, the model and equation (1) are undoubtedly simple. A problem immediately encountered concerns the values to be given to the parameters Φ^* and n_{WS} for a given metal. Another problem is to compile experimental values of ΔH for a large number of metal systems to enable an extensive study of the validity of equation (1) to be made. These problems will be discussed below.

It could be that making the two chemical potentials equal and smoothing out the discontinuity in the electron density are not independent effects. In that case a term proportional to $\Delta \Phi^* \Delta n_{WS}$ could be added to equation (1). In practice this mixing term is not found. It is evidently very small, and we have therefore neglected it for the sake of convenience.

The parameters Φ^* and n_{WS}

Until now Φ^* has been introduced as the chemical potential for electrons in an atomic cell. In more physical terms Φ^* could also be said to correspond to the position of the Fermi level in an absolute energy scale, e.g. with respect to vacuum. If a piece of metal one atomic cell in size still possessed the properties of a macroscopic piece of metal, it would be a straightforward matter to identify $\Delta \Phi^*$ with the contact-potential difference for the two pure metals A and B. In other words, $\Delta \Phi^*$ would be the same as $\Delta \Phi$, the difference between the work functions of the two metals. We shall try out this possibility, assuming that $\Phi^* = \Phi \pm 0.5$ V, a margin already present as an uncertainty in the experimental value for Φ. It would also have been possible to take the electronegativity as a measure of Φ^*, since the concepts of electron affinity and chemical potential for electrons approximately coincide. The difference is not very great, as can be seen in *fig. 5*, where Φ is

plotted against the electronegativity X in Pauling's scale for a number of metals. However, Φ seems preferable since it is a measurable physical quantity that can be expressed in well defined units.

How can a value be obtained for n_{WS}, the electron density at the boundary of the Wigner-Seitz cell of pure metals? For the alkali metals this is quite a simple matter. It may be assumed that the conduction electrons (one per atom) are fairly uniformly distributed in the bulk of the crystal. The density at the boundary of an atomic cell is then equal to the average density of one electron per atomic volume. The density n_{WS} for the simple polyvalent metals, the group of non-transition metals, can also be calculated by theoretical methods. It may be assumed that the charge distribution inside an Al crystal, for example, will not differ much from that obtained as a superposition of the charge distributions around free Al atoms in the crystal
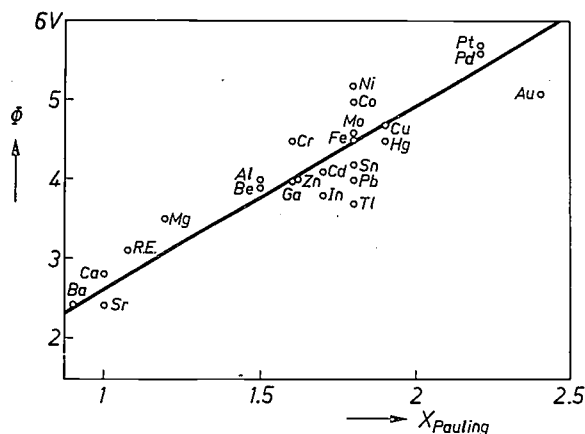
Fig. 5. Relation between the work function of a number of pure metals, Φ, and the electronegativity, X, in Linus Pauling's scale. The work function agrees best with the electronegativity scale Φ^*, which is the scale preferred for our approach.

structure formed by Al as a solid (f.c.c. lattice, molar volume 10 cm³). A reliable theoretical value can be calculated for the charge distribution around free atoms and consequently also for the superposition. Since the density at the boundary of the Wigner-Seitz cell is not everywhere the same, an average value is taken for n_{WS}.

It remains to see how the value of n_{WS} can be determined for transition metals. In their case it is no longer correct to treat the charge distribution in the metallic crystal as a superposition of the charge distribution of free atoms. It is necessary to resort to an empirical relation that connects the parameter n_{WS} with the compressibility of a metal. It seems reasonable to suppose that a metal in which the electron density between the

[9] See also A. R. Miedema, F. R. de Boer and P. F. de Chatel, J. Physics F (in the press). A more extensive treatment of the model will be found in: A. R. Miedema, J. less-common Met. 32, 115, 1973 and A. R. Miedema, J. Physics F (in the press).

atoms is low will be more easily compressible than a metal in which this density is high. *Fig. 6* gives the empirical relation found for all simple metals, for which n_{WS} can be theoretically calculated as outlined above. It is found that a linear relationship exists between n_{WS} and $(KV_m)^{-\frac{1}{4}}$, where K is the compressibility and V_m the molar volume. The graph also gives theoretical values for Cu and Fe (obtained from band structure calculations adjusted to be self-consistent) and a second value for Al, which was derived from experiments relating to small-angle X-ray scatter. Fig. 6 suggests that $(KV_m)^{-\frac{1}{4}}$ should be a good linear measure of n_{WS}. We shall take this as our starting point in determining values of n_{WS} for transition metals.

Demonstration of the applicability of the model

To check the validity of equation (1) we need quantitative information about ΔH for a large number of binary systems. To begin with we shall confine ourselves to 27 transition metals, i.e. eight metals from the 3d group, eight from the 4d group, eight from the 5d group and three 6d metals (see *fig. 7*). In the 5d metals the rare-earth metals are represented by La. The 27 metals in fig. 7 together form $27 \times 26/2$, i.e. 351 binary systems. Quantitative data on the heat of formation are available only for a few of these. We do, however, know the *sign* of ΔH in nearly all cases. For those binary systems in which intermetallic compounds are formed (or a phase with a crystal structure differing from that of the primary elements) that are found to be stable at low temperatures, it can be said with a fairly high degree of certainty that ΔH will be negative. However, in systems in which no intermetallic compounds are formed, and in which the mutual solubility is also small (e.g. less than 10 %), it may fairly safely be assumed that ΔH will be positive.

If equation (1) gives a correct description of the energy effect in alloys, then the difference in Φ^* for the two metals in binary systems where ΔH is negative must be relatively large compared with the difference in n_{WS}. If, on the other hand, ΔH is positive, the reverse should be the case. *Fig. 8* and *Table I* show that this reasoning does in fact hold good. Each point in fig. 8 represents a binary system characterized by $\Delta\Phi^*$ and Δn_{WS}. A minus sign indicates a system with negative ΔH, a plus sign a system with positive ΔH. It can be seen that a straight line may be drawn to correspond to $\Delta\Phi^*$ (in volts) $= 0.48\, n_{WS}$ (in density units), so that for nearly all minus signs we have $\Delta\Phi^*/\Delta n_{WS} > 0.48$, and for all plus signs $\Delta\Phi^*/\Delta n_{WS} < 0.48$. (One density unit — abbreviated sometimes to d.u. — corresponds approximately to 6×10^{22} electrons per cm^3.)

Apart from the groups of systems for which ΔH as
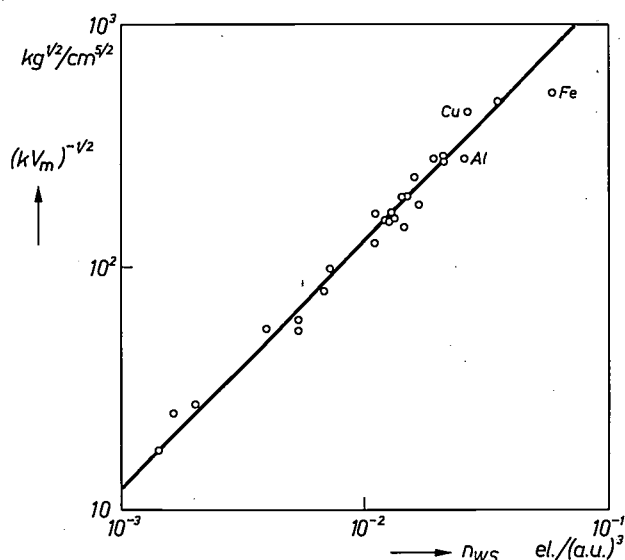
Fig. 6. Empirical relation between the product of the compressibility K and the molar volume V_m of pure metals and the theoretically calculated value for non-transition metals for the electron density at the boundary of the atomic cell, n_{WS}. Each point represents one metal; all non-transition metals are included. The three points for Cu, Fe and Al derive from another source (see text). The solid line corresponds to a linear relationship between $(KV_m)^{-\frac{1}{4}}$ and n_{WS}. The abbreviation a.u. stands for 'atomic unit'; 1 a.u. $= 0.572$ Å.

defined above has the plus or minus sign, there is of course another possibility to be considered. There will certainly be binary systems in which no compounds occur, and yet in which at least one of the two solubilities is greater than 10 at%. In these systems ΔH is neither distinctly positive nor distinctly negative. Our model indicates that ΔH must therefore be in the region of zero. This turns out to be the case: the circles representing these systems in fig. 8 are mainly concentrated near the origin. In view of the square terms $P(\Delta\Phi^*)^2$ and $Q(\Delta n_{WS})^2$ in equation (1), ΔH will not differ greatly from zero when both $\Delta\Phi^*$ and Δn_{WS} are small. A distinction is also made in fig. 8 between open and closed circles. The closed circles correspond to systems where the solubility at high temperature is greater than 10 % but decreases steeply at low temperature. In these systems ΔH is probably small but positive.

Sc 3.0	Ti 3.4	V 4.25	Cr 4.65	Mn 4.75	Fe 4.93	Co 5.1	Ni 5.35
Y 2.95	Zr 3.15	Nb 4.0	Mo 4.65	Tc 5.3	Ru 5.55	Rh 5.4	Pd 5.75
La 2.8	Hf 3.3	Ta 4.05	W 4.8	Re 5.6	Os 5.55	Ir 5.55	Pt 5.55
	Th 3.05	U 4.2	Pu 3.35				

Fig. 7. The 27 transition metals considered in this section, arranged in their order in the periodic system. The numbers give the values of Φ^* (in volts) that are relevant to the treatment of the magnetic properties of transition metals, discussed in Part II.

The result for alloys of transition metals provides immediately convincing confirmation of equation (1) for the heat of formation, and thus of the model as well. The slope of the straight line in fig. 8 gives the ratio of the two constants P and Q.

Table I. The electronegativity parameter Φ^* and the electron density at the boundary of the atomic cell, n_{WS}, for 27 transition metals. Φ^* is derived from the work function Φ (see text), and the electron density is derived from the compressibility and the molar volume of the pure metals. A unit of density (d.u.) corresponds to $(KV_m)^{-\frac{1}{2}} = 10^2$ kg$^{1/2}$ cm$^{-5/2}$, and is approximately equal to 6×10^{22} electrons per cm³.

Element	Φ^* (volts)	n_{WS} (d.u.)	Element	Φ^* (volts)	n_{WS} (d.u.)
Sc	3.0	1.7	La	2.8	1.0
Ti	3.4	3.2	Hf	3.3	2.9
V	4.25	4.4	Ta	4.05	4.3
Cr	4.65	5.2	W	4.8	5.9
Mn	4.75	5.1	Re	5.6	6.5
Fe	4.93	5.5	Os	5.55	6.7
Co	5.1	5.4	Ir	5.55	6.5
Ni	5.35	5.4	Pt	5.55	5.6
Y	2.95	1.4			
Zr	3.15	2.5	Th	3.05	1.7
Nb	4.0	4.0	U	4.2	4.1
Mo	4.65	5.5	Pu	3.35	2.2
Tc	5.3	5.9			
Ru	5.55	6.5			
Rh	5.4	5.7			
Pd	5.75	4.5			

For completeness we shall outline here the procedure by which the parameters Φ^* and n_{WS} were finally determined. A start was made by drawing a graph like the one in fig. 8, using $\Phi^* = \Phi$ and $n_{WS} \propto (KV_m)^{-\frac{1}{2}}$. At that moment the correlation between the sign of ΔH and $\Delta\Phi^*$ or Δn_{WS} was clearly present, but not so clearly as in the final fig. 8, since there were still a number of minus signs in the positive area and plus signs in the negative area. Next, Φ^* was allowed to differ slightly from Φ to see if this improved the correlation. A change in the value of Φ^* for a particular element causes a displacement of all points in fig. 8 that correspond to systems based on that element. By this process of trial and error the final, almost perfect correlation in fig. 8 was reached. For three metals (Mn, U, Fe) it was necessary to choose a value for n_{WS} that differed sligthy from the experimental values of $(KV_m)^{-\frac{1}{2}}$.

The final values for Φ^* still closely resemble the experimental values found for the work function, as can be seen in *fig. 9*, where the solid line corresponds to $\Phi^* = \Phi$. The open circles are the results obtained from the best correlation for fig. 8, plotted against recent experimental values reported by D. E. Eastman [10]. Relatively marked deviations between Φ^* and Φ are found for Ti, Zr, Hf (Φ^* much lower than Φ) and Mn.

Fig. 8 leads to the following simple rules for the metallurgical behaviour of transition metals in terms of the parameters Φ^* and n_{WS} from Table I:

[10] D. E. Eastman, Phys. Rev. B 2, 1, 1970.

A. When $\Delta\Phi^* > 0.48\,\Delta n_{WS}$, two metals readily alloy; this implies either that compounds exist in the system or that at least one of the solubilities is greater than 10% (240 cases with only one exception).

B. When $\Delta\Phi^* < 0.48\,\Delta n_{WS}$ for two transition metals, there are then no intermetallic compounds or ordered phases (105 cases, 4 exceptions).

These statements can also be reversed as follows:

A′. When intermetallic compounds (or ordered phases) occur in a binary system of transition metals, then $\Delta\Phi^* > 0.48\,\Delta n_{WS}$ (210 cases, 4 exceptions).

B′. When two transition metals do not alloy, i.e. when there are no compounds and the mutual solubility remains below 10 at%, then $\Delta\Phi^* < 0.48\,\Delta n_{WS}$ (65 cases, one exception).

It can be seen that the predictions obtainable with these rules, which in fact put the problem in a slightly different way from that found with the Hume-Rothery rule or the Darken and Gurry method discussed in the introduction, have an exceptionally high degree of probability. An interesting consequence of the rules is that it is now fairly straightforward to make predictions about the alloying behaviour of the missing 6d metals. It is sufficient to estimate, measure or define two parameters (as illustrated below for hydrogen), one approximately equal to the work function, and the other connected with the compressibility of the metal.

Although equation (1) and a list of Φ^* and n_{WS} values for the transition metals could obviously be used to give accurate rules for alloying behaviour, one or two questions remain. Is the description unambiguous? Is the freedom left in the choice of Φ^* so large that it will always be possible to find a good fit, even when the model on which equation (1) is based is not in itself correct? To learn something about the physical meaning of the Φ^* values obtained, an attempt has already been made in fig. 9 to see how far the relationship between Φ^* and Φ has been maintained; another possibility is that of extending the system of 27 transition metals by adding other metals. If we find that, without further adjustment of the Φ^* and n_{WS} values compiled, there is a clear correlation between the alloying behaviour of transition metals with that of the 'other metals', our confidence in the Φ^* values of Table I will be increased.

In *fig. 10a* the alloying of the 27 transition metals with Cu, Ag and Au is examined. If Φ^* and n_{WS} are assigned values close to the experimental values of Φ and $(KV_m)^{-\frac{1}{2}}$ for pure Cu, Ag and Au, a distinct correlation is again found between the sign of the heat of formation and the magnitude of $\Delta\Phi^*/\Delta n_{WS}$. In fig. 10a a straight line can easily be drawn separating the positive ΔH values from the negative ones. The slope of the line differs slightly from that in fig. 8: $Q/P = 0.255$

(eV/d.u.)² and $Q/P = 0.23$ (eV/d.u.)² respectively. In fig. 10a one minus sign is clearly in the wrong place (Cu-Mn lies in the positive region) and there is one plus sign wrong (Cu-Ta). There is no information available on the phase diagrams of Cu, Ag or Au with Tc or for Ag-Os. In these four cases the prediction on the grounds of fig. 10a is that there are no alloys (ΔH is positive).
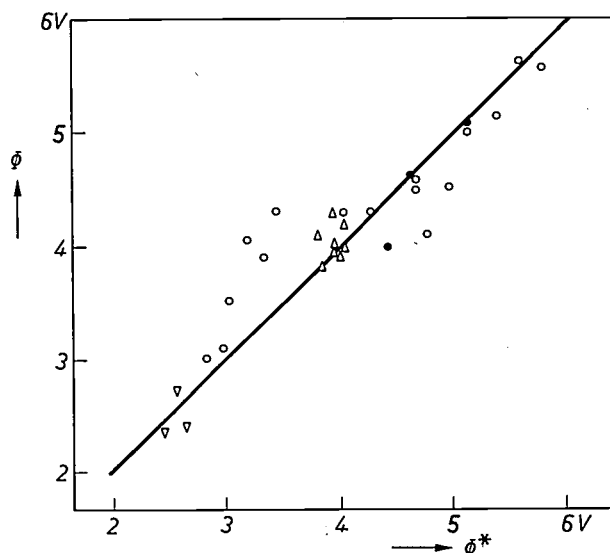
Fig. 9. Comparison between the parameter recommended here for the electronegativity of a metallic element, Φ^*, and experimental values for the work function Φ.

O transition metals	⎫ Experimental values given by D. E.
● Cu, Ag, Au	⎭ Eastman [10].
▽ Li, Ca, Sr	⎫ Experimental values given by
△ Al, Ga, In, Sn, Pb, Sb, Bi	⎬ H. B. Michaelson [11] and V. S.
	⎭ Fomenko [12].

Table II. Values for the electronegativity parameter Φ^* and the electron density at the boundary of the Wigner-Seitz cell for a number of non-transition metals. The experimental values of Φ are averages of values taken from review articles by H. B. Michaelson [11], V. S. Fomenko [12] and D. E. Eastman [10]. The density parameter is taken to be proportional to the experimental value of $(KV_m)^{-\frac{1}{2}}$, which can be calculated on the basis of tables given by K. A. Gschneidner [13].

Element	Φ^* (volts)	Φ_{exp} (volts)	n_{WS} (d.u.)	$10^{-2}(KV_m)^{-\frac{1}{2}}$ (kg$^{1/2}$ cm$^{-5/2}$)
Li	2.65	2.4	0.95	0.95
Ca	2.55	2.75	0.75	0.77
Sr	2.45	2.35	0.6	0.59
Al	4.0	4.0	2.7	2.72
Ga	3.9	3.95	2.2	2.21
In	3.8	3.8	1.6	1.63
Tl	3.95	3.9	1.4	1.45
Sn	4.0	4.2	1.9	1.84
Pb	3.9	4.0	1.6	1.55
Sb	3.75	4.1	1.5	1.46
Bi	3.9	4.3	1.2	1.23
Cu	4.6	4.65	3.8	4.3
Ag	4.4	4.0	3.2	3.2
Au	5.1	5.1	4.3	4.15
Mg	3.4	3.5	1.6	1.61
H	5.0		4.6	

Fig. 10b shows an analysis of the phase diagrams based on Li, Ca or Sr with 30 transition metals (now including Cu, Ag, or Au). There are now a fairly large number of phase diagrams about which no information is available (12 based on Li, 12 with Ca and 19 with Sr). Those that are available again show a good correlation with the $\Delta\Phi^*$ and Δn_{WS} values; as can be seen in *Table II* and fig. 10, the values for Φ^* and n_{WS} of these three metals are almost identical with the experimental values. The straight line in fig. 10b corresponds to

$Q/P = 0.27$ (eV/d.u.)^2, thus again to a value somewhat higher than the previous two. It can be seen that the extension of our original system of 27 metals to 33 does not require any further adjustment of the originally chosen Φ^* and n_{WS} values.

It is now possible to make a prediction about alloying or non-alloying for all binary systems based on Li, Ca and Sr about which no information is yet available. We can also do this for Ba and the heavier alkali metals Na, K, Rb and Cs, in which case the information al-
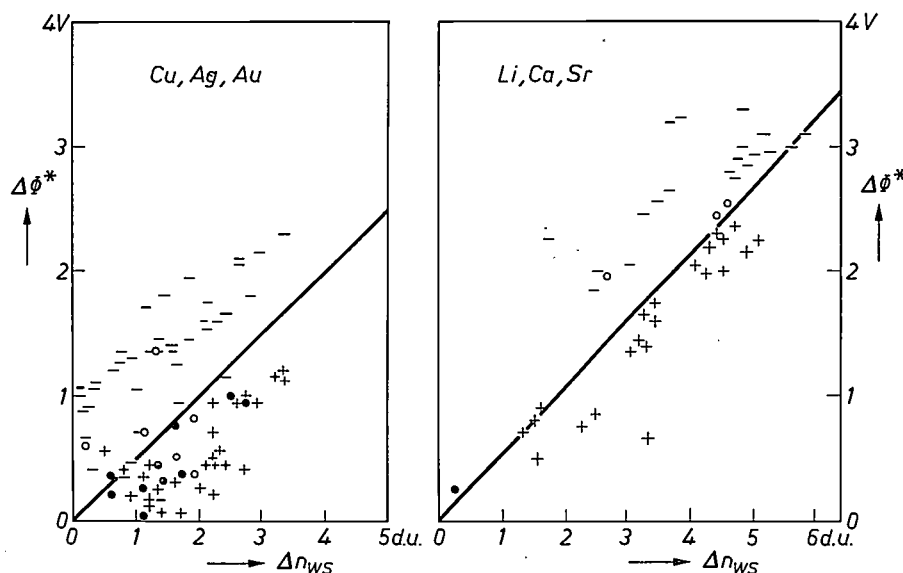
Fig. 10. The sign of ΔH for alloys of Cu, Ag, Au (*left*) and Li, Ca, Sr (*right*) with 27 transition metals.
— Stable compounds.
+ Eutectic with mutual solubility less than 10 at%.
O No compounds, solubility greater than 10 at%.
● No compounds, solubility greater than 10 at%, but ΔH is nevertheless positive. (ΔH was determined calorimetrically for low solubility at low temperature.)
-O- No compounds, solubility greater than 10 at%. Measured values are also given for ΔH, which has the negative sign.
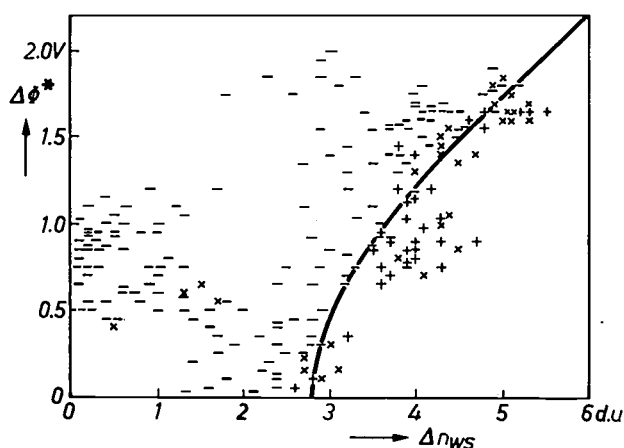
Fig. 11. Analysis of binary phase diagrams, based on one transition metal with Al, Ga, In, Tl, Sn, Pb, Sb or Bi. The case of no compounds occurring when one of the solubilities is greater than 10 at% is not found.
— Stable compounds.
+ No compounds and mutual solubility less than 10 at%.
× No information on the phase diagram available.

ready available on the alloying behaviour with the 30 transition metals becomes gradually smaller. The limited information available seems to suggest that a correlation like that of fig. 10b is fully maintained, but that the slope of the straight line continues to increase slightly in the series Na, K, Rb, Cs (the ratio P/Q changes and is not constant as it is between the 27 transition metals).

Fig. 11 analyses the alloying behaviour of the same 27 transition metals with the trivalent, tetravalent and pentavalent non-transition metals Al, Ga, In, Tl, Sn, Pb, Sb and Bi. The minus signs here indicate that there are compounds, the plus signs that there are no compounds and little mutual solubility, and the crosses that no phase diagram is known.

[11] H. B. Michaelson, J. appl. Phys. 21, 536, 1950.
[12] V. S. Fomenko, Handbook of thermo-ionic properties, Plenum Press, New York 1966.
[13] K. A. Gschneidner Jr., Solid State Physics 16, 275, 1964.

If again reasonable values are assumed for Φ^* (i.e. values that do not differ much from Φ), and using experimental values for $(KV_m)^{-\frac{1}{3}}$ as a measure of n_{WS}, a clear correlation is once again found between the heat of formation and $\Delta\Phi^*$, Δn_{WS}. There is an important difference from the foregoing, however. Equation (1) no longer appears to apply. It remains possible to draw a line separating the region with predominantly negative ΔH values from a region with purely positive ΔH values and unknown phase diagrams, but this line is no longer a straight one through the origin. The solid curve in fig. 11 relates to the simplest modification of equation (1) that corresponds to the results — the correlation found after the addition of a negative constant:

$$\Delta H = \mathrm{f}(c)\ [-Pe(\Delta\Phi^*)^2 + Q(\Delta n_{WS})^2 - R]. \quad (2)$$

The parabola in fig. 11 relates to $Q/P = 0.175$ $(\mathrm{eV/d.u.})^2$ and $R/P = 1.36\ \mathrm{eV}^2$. In this figure there are three minus signs in the positive region (Mn-Bi, Mn-Sb and Ir-Pb) and five plus signs in the negative region (Co-Pb, Ni-Pb, Re-Sn, Nb-Tl and Ni-Tl). The number of systems on which no information is available is 31. For 17 of these we can state positively whether they will or will not alloy; see Table III.

The extra negative contribution to the energy of the alloy, as found here for alloys of transition metals with non-transition metals whose conduction electrons have a distinct p character, is not known from the theory. However, the exceptionally large negative heat of solution of the rare-earth metals in molten tin (-60 kcal per gram atom) has earlier been considered a very surprising result.

It seems likely that the energy effect arises from more effective mixing (hybridization) of the wave functions of d electrons with p wave functions for the transition metals when they are associated with typical p metals in one alloy. What is surprising, however, is that this negative contribution should be approximately a constant that is virtually independent of which particular transition metal is alloyed with which particular p electron metal. The constant R represents a considerable contribution, as will later be shown (for dilute alloys R corresponds to a heat of solution of approximately 1 eV/at = 23 kcal/gram atom).

Visual presentation

It will meanwhile have become clear why the divalent metals Mg, Zn, Cd and Hg have not been considered along with the group of trivalent to pentavalent non-transition metals. Whether or not the energy term that corresponds to the constant R makes a contribution depends on whether or not these metals are alloyed with metals that are more strongly electronegative. When Mg is alloyed with, for example, Pd, Rh or Ni, it loses electrons and becomes a member of the group Cu, Ag, Au, Li, Na and K owing to the strongly ionic character

Table III. Predictions relating to the alloying of transition metals with trivalent or higher valency non-transition metals in cases for which no experimental information is yet available.

	Prediction: compounds	*Prediction:* no compounds and low solubility	
Definite	Pb-Hf Tl-Hf Bi-Hf Pu-Sb	Bi-Tc Bi-Ta Bi-Re Sb-W In-W Pb-Os	Tl-Mo Tl-Ru Tl-Cr Tl-W Tl-Re Tl-Ir Tl-Os
Probable	Sb-Re Sb-Tc Tl-Rh In-Re In-Tc Sn-Tc	Tl-V Tl-Tc Tl-Ta	

of the alloy. It is not possible to gain ionic energy while at the same time preserving the energy contribution connected with the presence of p electrons.

Let us examine this with the aid of *fig. 12*. Each metal is given a place in this figure determined by its Φ^* and n_{WS} values. If equation (1), without the extra constant R, were valid for Mg alloys, then it ought to be possible to separate the elements that do not alloy with Mg from the elements that do by drawing two symmetrical, straight lines, since $|\Delta\Phi^*| = a|\Delta n_{WS}|$ gives two straight lines. The metals in the upper and lower sectors should alloy well with Mg, whereas those in the left- and right-hand sectors should give a simple eutectic. Fig. 12 shows that the procedure gives excellent results; there is only one exception, the system Mg-Pu, in which compounds occur in spite of the almost identical Φ^*. The conclusion is that there might well be an extra energy term for Mg alloys, but that it is not large enough to stabilize compounds of Mg with Zr, Ti or Hf.

In *fig. 13* the procedure outlined for Mg is applied to three other cases. First it is applied to Au as a base metal to illustrate how the correlations found in this article can be used in practice. If the two lines corresponding to fig. 10a are drawn through Au as the origin, we see that one metal lies in the upper sector (Pd) and a large number in the lower sector. With all these metals Au forms intermetallic compounds or ordered superstructures, except with Ag, but a negative value for ΔH has in fact been measured calorimetrically in the system Ag-Au. For all metals in the right-hand sector ΔH should be positive. For Au-Pt and Au-Ni the value of ΔH is small, so that at high temperatures there may yet be full solubility. In the case of Au with Ru, Os, Ir, Re, etc. the mutual solubility is very small and there are no compounds. At the lower limit there is again some solubility in the case of Au-Mo (much more than in the case of Au-W), while Au again forms compounds with Cr.
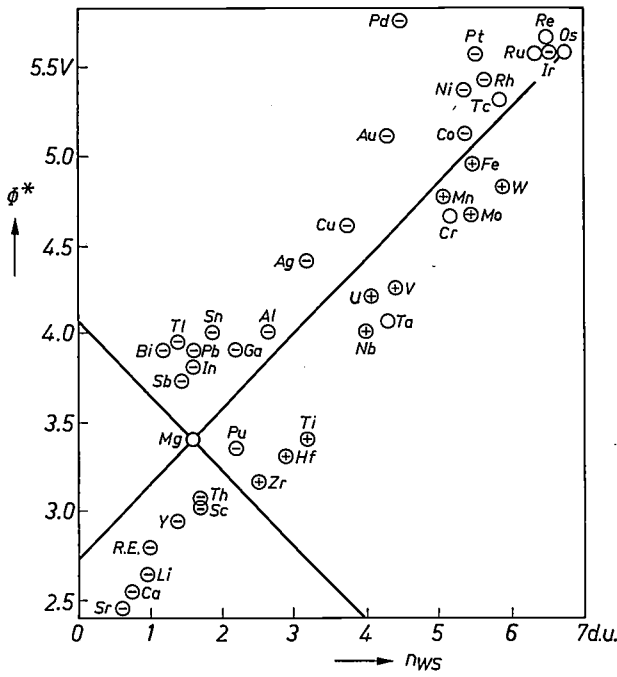
Fig. 12. Survey of the alloying behaviour of Mg. The minus sign indicates that intermetallic compounds with Mg exist, the plus sign means that no stable alloys are known. There is no information about Tc, Ru, Os, Re, Cr and Ta. When the values $\Phi^* = 3.4$ V and $n_{WS} = 1.6$ d.u. are assigned to Mg (i.e. virtually the experimental values), it is seen that ΔH is given by equation (1). The lines $|\Delta\Phi^*| = a|\Delta n_{WS}|$ separate the metals that alloy well with Mg (upper and lower sectors) from the metals that do not (left- and right-hand sectors).
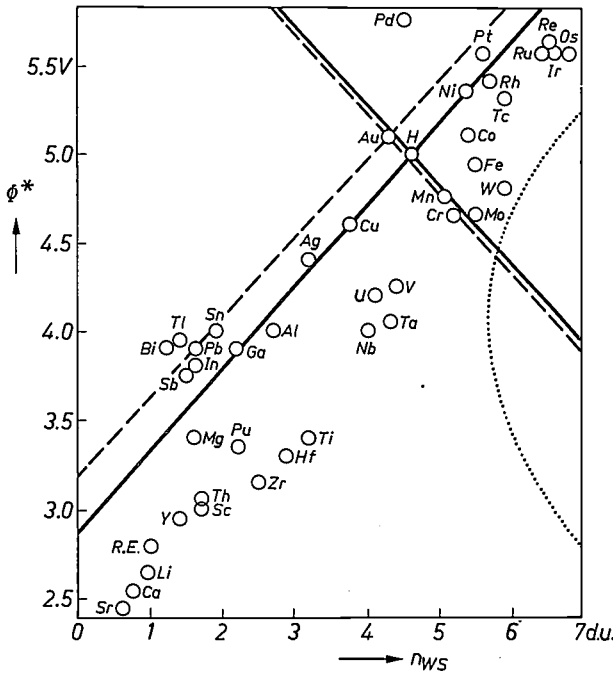
Fig. 13. Survey of the alloying behaviour of a large number of metals with respect to Au, H and Al. The lines through Au and H correspond to $|\Delta\Phi^*| = a|\Delta n_{WS}|$. The two lines that indicate where ΔH is zero for transition metals alloyed with Al form two parabolae (in accordance with equation 2), one of which falls outside the figure on the left.

——————— $\Delta H = 0$ with respect to hydrogen metal
— — — — — $\Delta H = 0$ with respect to Au
· · · · · · · · · $\Delta H = 0$ with respect to Al, for transition metals.

The only exception is in the left-hand sector: Au-Bi, which, like Au-Tl, ought not to form compounds but in fact does. The attractive feature of fig. 13 is that it shows at a glance the alloying behaviour of gold, which in itself is of course very well known.

A useful application is found in the metallization problem in semiconductor technology. In the finishing of semiconductor devices it is important that vapour-deposited gold layers that are in contact with other metals should not form intermetallic compounds with them, although some solubility is desirable to give adhesion. In fig. 13 it can be seen that Mo and also W would be a good choice here.

In the metallization problem it is also important to know which metals do not react strongly with Al. To demonstrate this visually for Al with transition metals we must draw not two straight lines, but two parabolae around Al as the centre-point. It is seen that Al alloys well with all transition metals that lie in the now interconnected upper and lower regions, and in accordance with the known facts (Al forms compounds with all the transition metals) there is not a single transition metal in the left- or right-hand sectors (the left-hand parabola falls outside fig. 13). The question that may now be asked is: with which transition metals are the compounds least stable? That is, which transition metals would be least reactive with Al? Fig. 13 provides the answer: in the first place W, then Mo, so that it has now been established that the combination Al-Mo-Au or Al-W-Au would be attractive in connection with the metallization problem.

Finally, it is worth pointing out that the two-parameter model can also clarify the reactivity of hydrogen with respect to the transition metals. The formation of a metallic hydride can be pictured as taking place in two steps. First, molecular hydrogen is converted into metallic hydrogen (at the expense of an energy amounting to some kcal per gram atom of H). After this, metallic hydrogen will alloy just like any other metal. Equation (1) would then be expected to apply for the heat of formation of the alloy, since this equation also gives a good description of the alloying of Li or Au. The question is: can values of n_{WS} and Φ^* be assigned to hydrogen metal such that the alloying behaviour of hydrogen fits in the picture outlined? The confirmatory answer is again given in fig. 13, were for hydrogen we take $\Phi^* = 5.0$ V and $n_{WS} = 4.6$ d.u. In accordance with the experimental data we find negative heats of formation for the alloying of H with Pd (in the upper sector) and with U, V, Nb, Ta, Pu, Zr, Hf, Ti, Th, Sc, Y and the metals of the rare-earth group in the lower sector. The borderline cases Mn and Cr (heat of formation almost zero) also agree with the model, the only exception being Pt. An important consequence of the

analysis presented here is that, in the stable hydrides, hydrogen is the electropositive element only in the system Pd-H. For all other binary systems there can be little doubt that hydrogen acts as the more electronegative metal, which accepts electrons.

To summarize the main arguments of this first part of the article, we can say first that the two main contributions to the heat of formation of alloys of two metals are made by a negative term, connected with the difference in the electronegativity of the two metals, and a positive term that corresponds to the difference in electron density at the boundary of the atomic cell in these metals.

The proof that these contributions are not only present for alloys of two transition metals, but are also sufficient for determining whether two metals will alloy or not, is obtained from the ability to formulate rules for the alloying behaviour that allow predictions to be made with an exceptionally high degree of certainty. For alloys of transition metals with polyvalent non-transition metals, a third contribution to the alloying energy has been found that does not depend on which transition metal is alloyed with which p metal. In all

this the size of the atoms plays no part. It only becomes important when it is a question of determining whether two metals can form solid solutions; it is not relevant to the question of whether stable alloys exist.

The recommended values for the electronegativity of metallic elements correspond closely to the experimental values found for the work function of pure metals.

Summary. Reliable rules have long been sought for the alloying behaviour of metals. There is the qualitative rule that states that the more two metals differ in electronegativity (a rather vague concept), the greater the heat of formation of the alloy and thus its stability. There is also the Hume-Rothery rule, which states that two metals differing by more than 15% in their atomic radius will not form solid solutions. This rule can only be used to predict poor solubility (90% reliability), it will not predict good solubility. In this article a simple atomic model is presented, which although empirical like the other rules nevertheless has a clear physical basis and predicts the alloying behaviour of transition metals correctly in 98% of cases. In this model there are two contributions that determine the heat of formation: a negative one, connected with the difference in electronegativity — this quantity must be taken as proportional to the experimental value of the work function — and a positive one, corresponding to the difference in electron density at the boundary of the atomic cells of the metals. For alloys of transition metals with polyvalent non-transition metals there is also a third contribution to the heat of formation, which is independent of which transition metal is alloyed with which p metal. None of this is affected by the atomic size; this is only of significance in deciding whether two metals are able to form solid solutions, and not whether stable alloys exist.
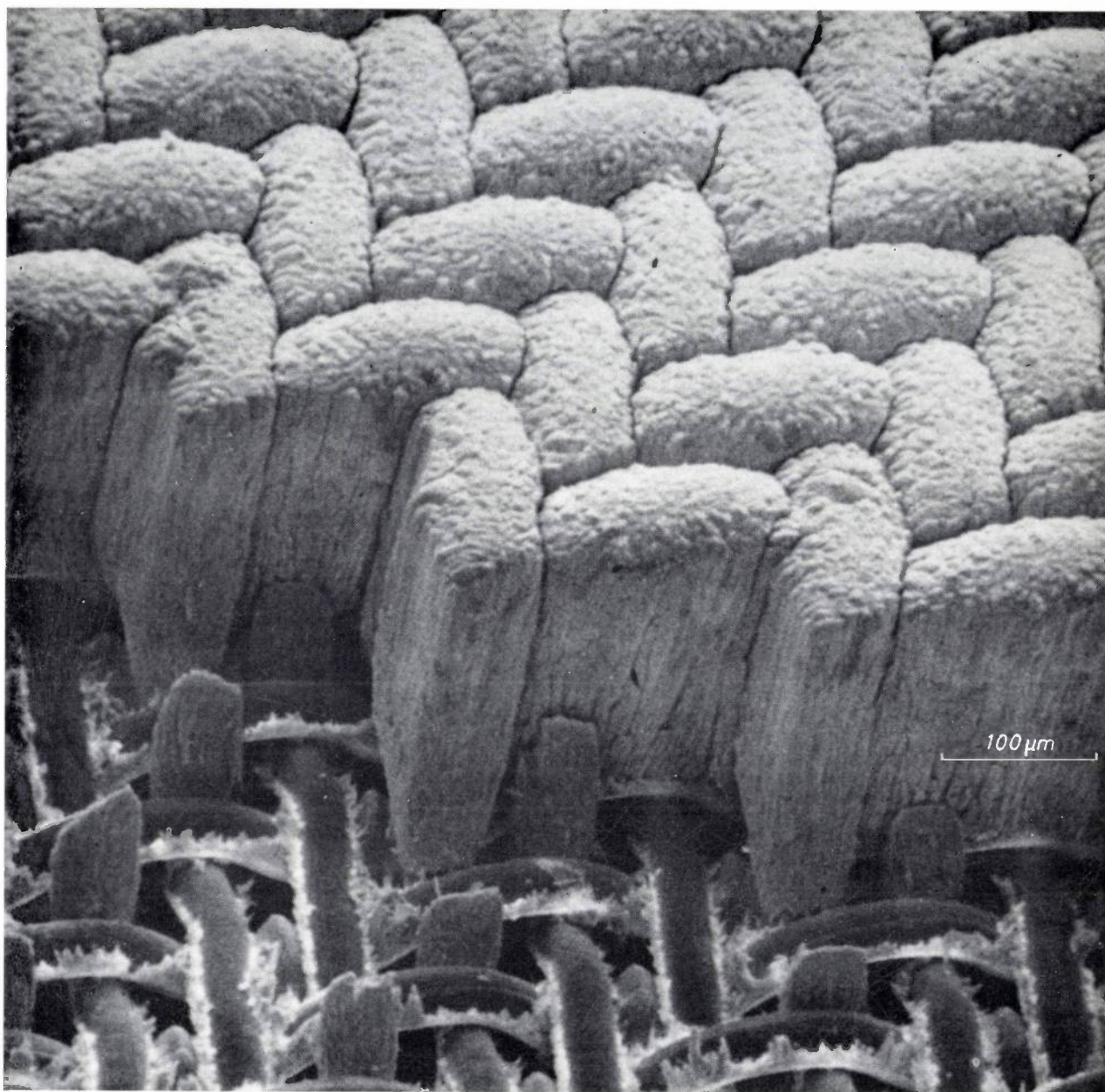
Luminescent screen with 'pillar structure'

A luminescent screen that converts X-radiation into visible light must have an appreciable thickness (200 μm or more) if it is to absorb the incident radiation sufficiently. A thicker layer, however, means a poorer resolution, owing to lateral scattering of the generated light. This can be avoided to some extent by using a screen structure that may be likened to a layer of 'pillars' standing vertically on the substrate of the screen. The preference for propagation of the light in the longitudinal direction of the pillars is based on repeated reflection of the light from the smooth side faces. With some substances (e.g. CsI) a structure of this type is obtained by deposition from the vapour phase on to the substrate, but in that case there is fairly strong optical contact between the pillars. This is usually avoided by creating light barriers in a subsequent thermal treatment of the deposited layer, causing narrow fissures to form between clusters of pillars.

A method has been devised by A. L. N. Stevels and A. D. M. de Pauw at Philips Research Laboratories in which these structures are produced in a controlled way during the actual vapour-deposition process. In this method the substrate is previously given the desired pattern in some way or another. The pattern can be made regular in form, and the dimensions of the elements can be freely chosen within certain limits. The photograph (made by J. L. C. Daams with a scanning electron microscope) shows one of the experimental forms, in which the substrate is a very fine mesh of molybdenum-wire gauze. The vapour-deposited layer is CsI with a thickness of about 200 μm. Layer growth begins on the protruding parts of the gauze, and continues until completely separated clusters of pillars have formed. Even if the layer is made considerably thicker, it still retains the pattern of the substrate.

162

Philips tech. Rev. 33, 162-176, 1973, No. 6

Attitude control for the Netherlands astronomical satellite (ANS)

P. van Otterloo

This second article on the Netherlands astronomical satellite (ANS) deals with the attitude-control system. Astronomers who wish to make measurements on what are virtually point sources require their telescopes to be aimed extremely accurately. To satisfy those requirements the satellite has been equipped with a control system whose accuracy and flexibility put ANS in the forefront of modern space technique. The secret lies in the the digital processing of the data, by software in an onboard computer instead of using specific hardware.

Purpose of attitude control and how it works

In most astronomical investigations, whether from the Earth or from space, the accurate and constant aiming or 'pointing' of the observation equipment at the object under study is a first prerequisite. This is clearly reflected in the equipment of the astronomical satellite ANS ('*Astronomische Nederlandse Satelliet*'). The introductory article [1] on the ANS project characterizes the satellite as a relatively comprehensive package of astronomical observation equipment with a refined, locally operating and flexible service system, whose main task is the automatic control of attitude for pointing the astronomical instruments. It was stated in the earlier article that the maximum total mass of the satellite would be about 130 kg, and the construction therefore was to be kept particularly simple. For this reason the sensors for the attitude control, the deployed solar panels and the astronomical equipment are rigidly connected with the frame of the satellite, without any possibility of relative movement. Mechanical restrictions of this type do not of course make attitude control and operational aspects any easier. Another restriction is that the mass of the attitude-control equipment should be no more than the 15.5 kg, permitted in the 'mass balance' schedule for the satellite.

The control follows a sequence consisting of ten operating modes. *Fig. I* shows a number of these modes, which are completed after separation of the last launching stage. The first is the 'despin' mode, which takes place in two steps, to eliminate a rapid rotation (180 rpm) originating from the launching vehicle. Next the satellite takes up a position such that

its solar panels are oriented perpendicular to the line connecting it to the Sun. This is followed by a mode in which the sky is scanned in a circular arc in a search for an object to be observed. In the scanning mode the satellite slowly rotates about the axis constantly directed towards the Sun. Finally the attitude-control system ensures that the instruments are kept permanently pointed at the object (fine pointing).

The attitude control is effected by means of an automatic control loop, in which a number of sensors regularly measure the attitude of satellite and observation equipment and report this to an onboard digital computer [2]. After comparison of the measured attitude data with the desired values (or setpoints), the computer sets actuators in operation to carry out any corrections required [3]. The purpose of the attitude-control system is in fact twofold. On one hand it has to point the astronomical instruments in certain directions or at certain stars on receipt of commands from the onboard computer and to keep the instruments pointed; and on the other hand it has to allow the satellite to take up a position such that both of its solar panels continuously receive the maximum amount of sunlight, since otherwise the power supply could become insufficient.

The procedure of pointing the satellite at the Sun is preceded by a 'Sun-acquisition' phase, in which the satellite looks for the Sun. During Sun acquisition the

Ir P. van Otterloo is with Philips Research Laboratories, Eindhoven, and is Local Project Manager for ANS.

[1] W. Bloemendal and C. Kramer, The Netherlands astronomical satellite (ANS), Philips tech. Rev. 33, 117-129, 1973 (No. 5).

[2] G. J. A. Arink, The onboard computer for the Netherlands astronomical satellite (ANS), to be published in a forthcoming issue of this journal.

[3] Separate articles on the sensors and actuators will be published later in this journal.
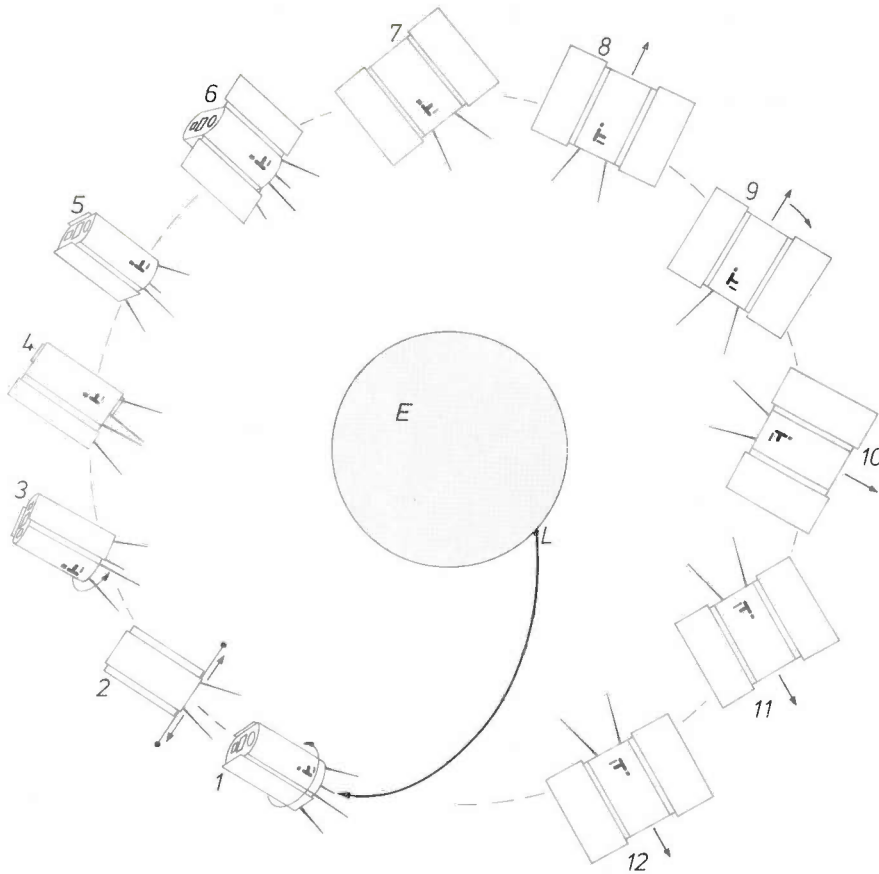
Fig. 1. The successive situations or operating modes of the satellite in a normal mission, as they would appear from the Sun. *E* Earth. *L* launch site. *1* The satellite is first in a state of rapid rotation, caused by the spin stabilization of the launcher, and describes a *nutation* movement due to the separation from the last stage. *2-4* The 'despin' and 'detumbling' modes. *5-7* The front of the satellite is turned towards the Sun; the two panels with solar cells are now deployed. *8-12* Rotation of the satellite about the axis pointing towards the Sun to search for stellar objects; the astronomical instruments 'look' in the direction of the arrow.

electrical power is supplied by a nickel-cadmium battery. The battery is also used when the satellite for one reason or another, such as a slight orbit-injection error, has to spend some time in the Earth's shadow at the beginning of the mission. This situation is signalled by a group of photosensitive sensors. The control devices on board then take measures to minimize the consumption of power.

An important subsidiary task of the attitude-control system is to supply attitude data, which are continuously picked up by the various sensors, for storage in the memory units of the onboard computer. After transmission to Earth, these data can be taken into account when the results obtained with the astronomical instruments are evaluated.

In the following discussion of ANS it will be convenient to assume that it has a fixed coordinate system (*fig. 2*). When folded out the solar panels lie in the *x,y* plane. The positive *z*-axis is perpendicular to the front plane of the satellite, and after Sun acquisition it points accurately in the direction of the centre of the Sun. The optical axis of the Cassegrain telescope carried on board [1] acts as the *x*-axis. All other observation instruments, such as the proportional counters for X-radiation and all attitude-control sensors, are aligned in relation to the Cassegrain telescope to within a tolerance of no more than three minutes of arc. The equipment is fixed to the frame of the satellite in such a way that any slight deformations of the frame will cause the least misalignment. The final calibration of the alignment takes place by adjustment of the computer programs when the satellite has been put into orbit (in-orbit checkout).
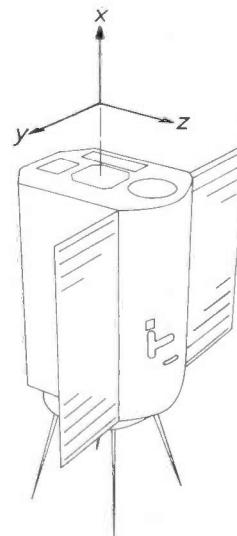
Fig. 2. Sketch of the satellite showing the coordinate system used. The *z*-axis is perpendicular to the plane of the deployed solar panels. The *x*-axis points in the direction in which the astronomical instruments can make their observations.

The attitude control chosen for ANS is an 'active' three-axis system of the coarse-fine type. This means that there are sensors and actuators on board that are capable of continuously detecting and correcting changes in satellite attitude in *three* dimensions. The adjustment is coarse in relation to the Earth's horizon, coarse and fine in relation to the Sun-satellite line, and fine in relation to the line between the satellite and the object of observation. For these functions there are altogether four types of sensor and three types of actuator on board.

'Passive' attitude-control systems also exist, such as spin stabilization and gravity-gradient stabilization systems. In the second system the stabilizing force is derived from the difference in the gravitational forces exercised by the Earth on two heavy masses located some distance apart in the satellite. This implies alignment in relation to the local vertical, which, since there is then no orientation towards the Sun, makes the system unsuitable for ANS.

For spin stabilization one axis of the satellite is kept in a fixed attitude in space; a flywheel in the satellite [4] (or the satellite itself) rotates at a speed of about a thousand (or a hundred if the satellite itself) revolutions per minute about the appropriate axis. A system of this type would make it impossible for ANS to carry out the observations, which have to be made in all directions. The direction of the axis of rotation would of course be difficult to change, and in lateral observations the viewing direction would change much too quickly.

As mentioned earlier, the entire attitude-control system is designed as a digital system comprising an onboard computer as the principal controller (*fig. 3*). This digitalization makes it very easy to standardize the interface between any two neighbouring elements in the control loop. An even more important point is that it gives greater operational flexibility. Since the data processing is performed by software stored in the computer memory units, it is much easier to make modifications or adjustments in the processing than it would be if the processing methods were based on analog
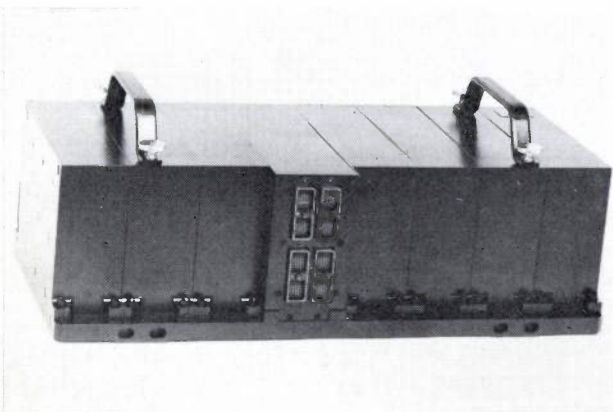
Fig. 3. Onboard computer with storage. The part with the connectors is the central processor. The other modules form seven independent storage units.

circuits. The total storage capacity for control and measurement data is 28 000 words of 16 bits. The possibility of introducing changes from the Earth into the available control algorithms has the great advantage that even if a fault should occur during the flight, the satellite can still be optimally used simply by correcting the software.

In the following sections we shall first take a closer look at the design specifications for the attitude-control system; as we shall see later, these specifications were primarily determined by the astronomers' requirements. A description will then be given of the various elements used in the control system, principally the sensors and actuators, and also of the operating modes of the satellite after separation from the last stage of the launcher. Finally there will be a brief description of the way in which the attitude-control system has been checked in the laboratory.

Design and main elements of the control system

As stated, the performance required of the attitude-control system is very closely related to the measurements that ANS has to carry out. The spectrometer behind the telescope for investigating the ultraviolet radiation from young stars [1] has an effective field of view of 2.5×2.5 minutes of arc. The image of a star is of course very nearly a point; the diameter corresponds to an angle of less than 10 seconds of arc. During the measurement the optical axis of the telescope and the line between the telescope and the star should therefore not differ in direction by more than $1\frac{1}{4}$ minutes of arc.

Before the telescope can be aimed at a star and kept tracking it, the star must of course first be detected. Unfortunately the radiation from the stars under investigation (with brightnesses up to the 10th magnitude) may often be far from strong enough to provide a useful reference for pointing. To get around these problems ANS will work with a 'pattern-recognition' system, enabling the attitude-control system to recognize an object from two brighter guide stars in the immediate neighbourhood. The quality of the fine pointing will then be independent of the brightness of the object under investigation, which is a great advantage in the measurement of the many extremely weak radiation sources which the astronomers have in their programme. To make sure that there are sufficient guide stars over the whole celestial sphere, the system will be made sensitive enough for stars with a brightness up to the 8th magnitude to be used as guides. Another special feature of ANS is that the pattern recognition will be made through the telescope used for the astronomical measurements. The mechanism for this purpose is incorporated in the tele-

scope, which means that there are then no alignment problems.

Since the X-ray detectors are aligned with respect to the telescope, the detection and tracking of the objects under investigation can take place in the same way with two guide stars via the telescope. An extra feature in the design of the attitude-control system is that a source of sufficiently *hard* X-radiation can be tracked directly by means of the output signal of the X-ray detector, and not by making use of guide stars. The X-ray detector is fitted with a collimator that defines a sufficiently sharp direction for pointing.

During the six months in which the satellite is operational the z-axis of ANS rotates through an angle of 180 degrees. Because of this the possibility of observing an object in the celestial sphere is limited to one particular day, which is the day on which the plane perpendicular to the z-axis contains the object in question. Furthermore, during any revolution around the Earth no more than 30 minutes of astronomical observations can be made on the same object. Otherwise sunlight reflected from the Earth would penetrate the telescope; scattered light upsets both attitude control and observations as soon as the angle between the local vertical and the optical axis of the telescope is more than 75°. A time of 30 minutes is too short for certain observations, such as those for determining the frequency of intensity fluctuations in the X-ray emission from pulsars [1]. Observations of this type have to be repeated in a number of successive revolutions, and this is only feasible if the attitude-control system compensates the simultaneous slight shift of the z-axis, so that the plane perpendicular to it continues to contain the source. A refinement has been built into the system for this purpose; it allows stabilized misalignments of the positive z-axis to be programmed from the Earth, up to a maximum of 0.7° (offset facility). This enables observations to be made on the same object during some twenty successive orbits. The specified average accuracy with which the z-axis is to be pointed at the Sun is normally 0.01°.

The actuators

Three types of actuator have been chosen for making corrections to the attitude and movements of the satellite. The initial spin due to the launcher will be stopped by a 'yo-yo'. This actuator consists in principle of two auxiliary masses fixed to the satellite, which fly out shortly after separation of the last launcher stage, each mass being attached to a taut release line a few metres long (*fig. 4*). The auxiliary masses exert an opposing torque until the satellite has lost nearly all its spin. When the release lines come into the radial position they automatically become detached (if this did

not happen the angular momentum carried by the auxiliary masses would be returned to the satellite). The yo-yo is thus used once only, and therefore more actuators are needed.
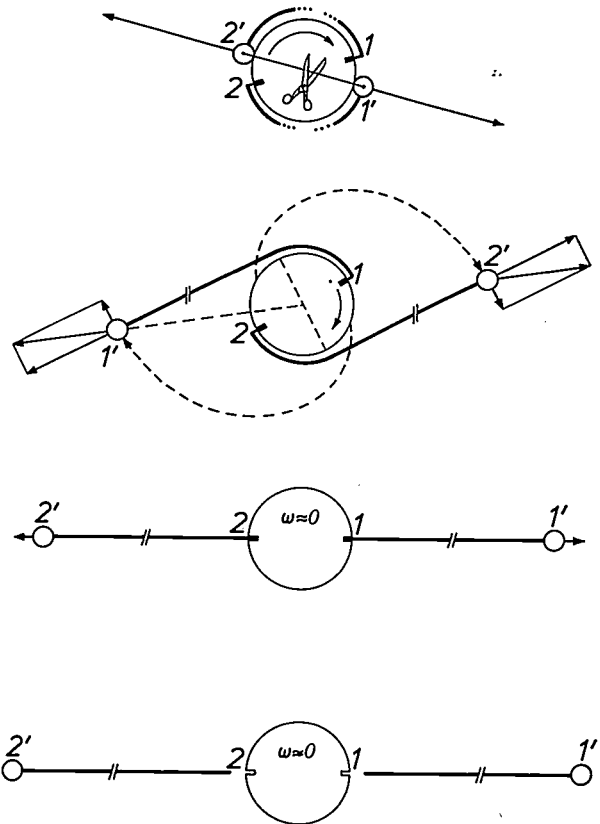
Fig. 4. Diagram of the 'yo-yo' despin system, which reduces the initial spin rate of the satellite (180 rpm) to about 2 or 3 rpm. The system is carried on a ring located below the base plate of the satellite frame; this ensures that the release lines function freely. The yo-yo comes into operation as soon as a 'guillotine' mechanism cuts a connecting line that holds the yo-yo masses in place. The guillotine is controlled by a time switch that starts when the satellite separates from the last stage of the launcher. The masses and the length of the release lines are given values such that the initial spin has just reached zero ($\omega = 0$) when the lines have completely unwound.

The second actuator is a magnetic 'torquer', consisting of three air-cored coils (*fig. 5*), mounted at right angles to each other in the satellite. These coils can be energized independently of one another, under the supervision of the attitude-control logic, which will be discussed in the section dealing with the complete control system. The interaction between the coils and the local magnetic field of the Earth makes it possible to exert a variable torque in such a way as to change the angular momentum of the satellite. This facility is subject to limitations, however, since no torque can of course be generated to provide a rotation with the

[4] See for example: J. P. Reinhoudt, A flywheel for stabilizing space vehicles, Philips tech. Rev. 30, 2-6, 1969.

local field vector as the axis of rotation, and moreover, the accuracy achieved with a magnetic actuator of this type, a simple on-off control with dead band, is unfortunately not very high, which explains the presence of the third actuator.

The third actuator consists of a system of three orthogonal reaction wheels (*fig. 6*), which act as angular-momentum exchangers. As the name suggests, a reaction wheel only exchanges angular momentum between the satellite and itself (the angular momentum of the complete satellite remains unchanged). Since these reaction wheels can be very accurately and easily controlled by the onboard computer, using an electric

Fig. 5. One of the coils of the magnetic 'torquer'. There are three such coils at the base of the satellite, each perpendicular to one of the coordinate axes (fig. 2). The coils are orthocyclically wound [5] with aluminium wire and embedded in a thermoplastic material. The power required is 2 W. Each coil delivers a maximum torque of 5×10^{-4} Nm.
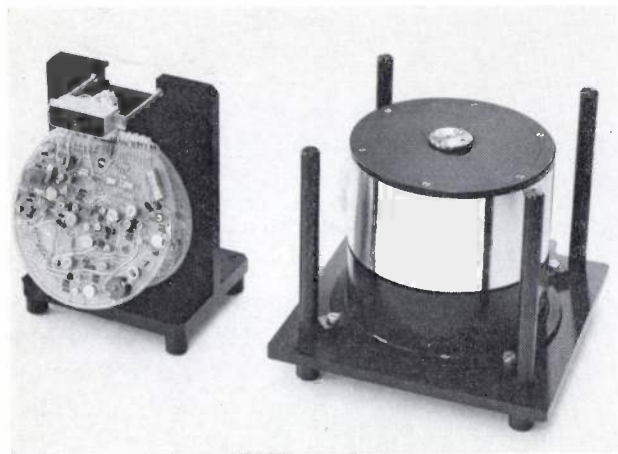
Fig. 6. One of the three reaction wheels. The shafts of the reaction wheels are parallel to the three coordinate axes (fig. 2). Each wheel can deliver a positive or negative torque, with a total of 23 different values (including zero). The maximum value is 1.5 Ncm. The maximum power taken is 4 W per wheel. One of the control-logic units is shown on the left. Each unit consists of a set of printed-circuit boards and is mounted inside the black cylinder below the wheel.

servomotor drive, the exchange of angular momentum can thus be made to take place with sufficient accuracy.

The attitude-control logic is designed to ensure that the magnetic actuator interacts in the correct way with the reaction wheels. If, for example, a reaction wheel takes up so much angular momentum that it approaches its maximum permissible speed of revolution, the wheel itself will signal this state of affairs to the attitude-control logic, which will instruct the magnetic actuator to exert an appropriate torque on the satellite as a whole so as to remove the necessity for the excessively high speed of the reaction wheel. The wheel will then start to turn more slowly, i.e. it is 'discharged' by the magnetic actuator.

Although it might appear at first sight to be a useful technique, actuators based on expulsion of mass are not used in ANS. The primary objection to such actuators for an astronomical satellite is that they produce clouds of gas, which can adversely affect the operation of the astronomical instruments.

The solar sensors

The solar sensors consist basically of one or more photosensitive silicon devices (solar cells) housed in such a way as to give a well defined field of view. In the first stage of Sun acquisition the satellite will tumble erratically through space. For this reason a sensor is needed that is capable of scanning rapidly the *entire* celestial sphere (*fig. 7*). On the other hand the final pointing of the positive z-axis at the centre of the Sun must be very accurate. One sensor is not capable of meeting both these requirements, and therefore in addition to the 'coarse' solar sensor just mentioned, there are two others included in the design: an intermediate solar sensor (*fig. 8*; maximum pointing error 1°) and a fine solar sensor (*fig. 9*; maximum error 0.01°). These sensors are required to measure the sign and magnitude of rotations about both the x-axis and the y-axis, since any directional error of the z-axis is equivalent to the superposition of rotations about the two other axes. The easiest solution is to duplicate the two sensors, one for the x-axis and the other for the y-axis.

The high directional sensitivity of the fine solar sensor is only of significance if the alignment of the sensor remains within the same narrow limit as 0.01° in relation to the astronomical observation instruments. This sensor is therefore rigidly fixed to a base plate of the Cassegrain telescope used for the observations.

The housings that define the field of view of each of the solar sensors range from simple masks, which allow a conical beam of light to fall on a solar cell, to complete small telescopes with cylindrical mirrors, which

form a linear image of the Sun somewhere on a row of 200 photosensitive elements. The position, and also any shift in the linear image with respect to these elements, which operate as a shift register, then give an extremely sensitive directional indication for the on-board computer. The linear image covers 50 elements, i.e. one quarter of the total field of view of 2°, the diameter of the Sun corresponding to an angle of 0.5°.

The horizon sensor

After the fine pointing of the z-axis the next step is to point the observation instruments at an object to be studied, in other words to point the x-axis. Clearly, the Sun cannot serve as a reference for this attitude control 'around' the z-axis. The obvious thing to do is to bring the celestial body closest to hand, the Earth, into the control system. The horizon sensor (*fig. 10*) measures the direction of the x-axis with respect to two connecting lines between satellite and horizon; these are the connecting lines lying in the x,y plane. This sensor is only sensitive in the wavelength region in which the Earth most resembles a uniformly emitting disc (the infrared around 15 μm). The operation of the sensor is little affected by geographic latitude or strongly alternating conditions such as night and day. The sensor detects the transition between Earth and space (and vice versa) by a scanning movement, and gives the desired direction of the x-axis to within an accuracy of about 0.3°. This means of course that the direction of the x-axis with respect to the local vertical can also be calculated.

The ability to observe a celestial object implies that the x-axis must have a specific direction in *space*. It will be clear from fig. 10b that this viewing direction in space for an x-axis with a fixed attitude angle α in relation to the line between satellite and horizon will only occur when the satellite in its orbit around the Earth passes a particular point with an angle of latitude δ. The expected uncertainty with regard to the exact value of δ is unfortunately so great that the correct point in the orbit for the observation would be communicated to the computer too inaccurately. This means that fine pointing at the object, and in particular the tracking of the object, would fail. For this reason the horizon sensor is used only for *coarse* pointing of the observation instruments. The fine adjustment is made with another type of sensor, which will now be discussed.

The star sensor

For fine pointing at a stellar object and for star tracking, the satellite uses a star sensor (*fig. 11* and

[5] W. L. L. Lenders, The orthocyclic method of coil winding, Philips tech. Rev. **23**, 365-379, 1961/62.
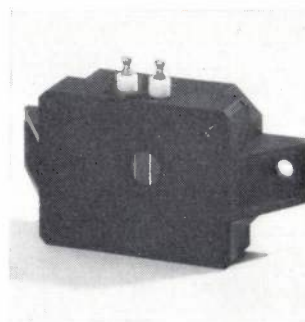
Fig. 7. One of the silicon devices forming part of the coarse solar sensor, mounted in its housing, which also serves as a mask. One such device is fitted on each of the six sides of the satellite. The fields of view are cones with an apex angle of 120 degrees and overlap one another.
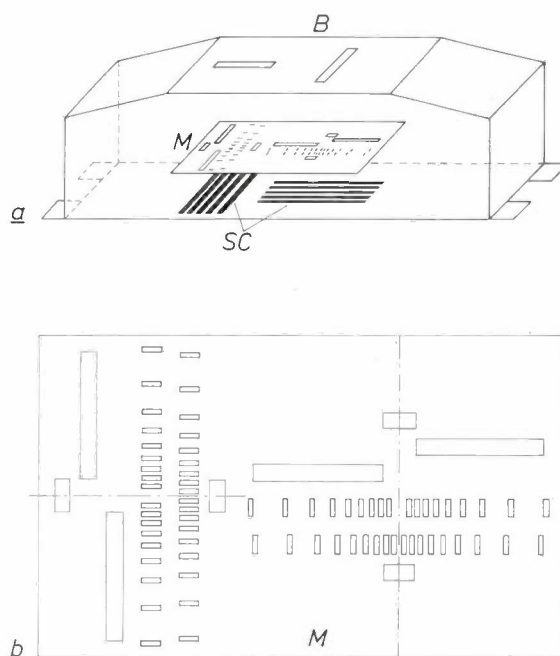
Fig. 8. *a*) Schematic representation of the intermediate solar sensor. A box *B* with two slits in it contains two series of five solar cells *SC*, each diffused into a 60 × 40 mm silicon chip, over which a mask *M* with a pattern of slits is fitted. Both series of solar cells have a field of view of 72°; one is in the x,z-plane and the other in the y,z-plane. The pattern of slits is laid out so as to make it possible to determine from the signals the direction and speed of rotation about the x- and the y-axis. The total power taken by the sensor is 0.15 W. *b*) The mask.
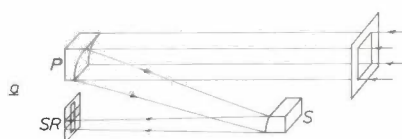
Fig. 9. The fine solar sensor. The detector *SR* is mounted in the focal plane of a small cylindrical telescope (*P* primary mirror, *S* secondary mirror); it contains a shift register composed of a row of 200 photosensitive bistable circuits (*P*-channel MOS transistors). Field of view 2°, power required 0.2 W. There are two of these sensors, one for the x,z-plane and one for the y,z-plane. The position of the linear image of the Sun on the shift register can be determined very accurately.

Fig. 10. *a*) The horizon sensor. This is located on the side of the satellite facing away from the Sun. The plane mirror *M* rotates at an angle of 45° about the *z*-axis, so that the sensor has a field of view of 360° in the *x*,*y*-plane. The germanium lens *L* concentrates the received radiation on the temperature-dependent resistor *R*. At an abrupt change in *R* the position of the mirror is read out. *Mo* motor. Total power required 1 W.
b) Pointing the *x*-axis with the aid of the horizon sensor. The *z*-axis (perpendicular to the plane of the drawing) points towards the Sun. The horizon sensor detects the horizon in two directions, and the onboard computer calculates the angle *γ* between the *x*-axis and the local vertical from the angles *α* and *β*. The latitude angle *δ* establishes the position of the satellite with respect to the Earth's equator *Eq*.

fig. 12) which looks through the Cassegrain telescope of the ultraviolet observation system [1]. The star sensor is a camera tube of the image-dissector type. It consists of an electron-optical image scanner (using a photo-cathode) followed by a diaphragm behind which is located a multiplier section. The tube combines the required sensitivity with an absence of any significant storage effect. This feature is necessary, otherwise slight movements of the satellite would cause a blurred picture. Radiation entering the telescope at a small angle to the optical axis is imaged on the photocathode of the camera tube. (The radiation entering parallel to the optical axis is the radiation investigated in the astronomical measurements.) To detect and track an object the onboard computer compares the information received from the camera tube with its own data on two guide stars, and from this calculates the control signal that is needed. The attitude-control system then manoeuvres the satellite in such a way that the image of one of the guide stars is kept constantly at the same place on the photocathode. If this place is well chosen, the radiation from the object under observation will then at the same time enter accurately along the optical axis of the observation instrument and the associated detector. It is to be expected that in this way the directional deviation will remain below the maximum permissible value of 1 minute of arc.

The magnetometer

Besides the three kinds of optical sensor the satellite has another sensor that measures the components of the Earth's local magnetic field; these data are used in the attitude-control logic for calculating the energizing of the magnetic torquing coils. Measurement of the Earth's field is needed for the detumbling of the satellite
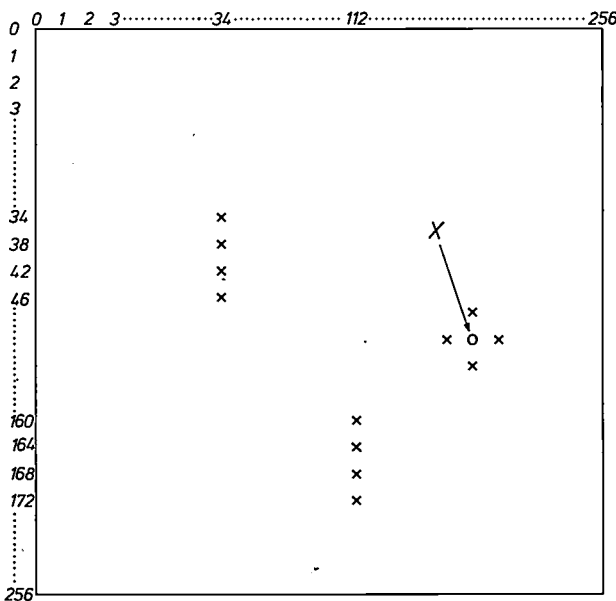
Fig. 11. The detector part of the star sensor is an image-dissector tube. A picture is projected on the photocathode of the tube, which is divided into 256 × 256 resolution elements. The onboard computer indicates a group of four resolution elements in a particular column, which is continuously scanned. If this group delivers a signal, as a result of the detection of a star, a second group of four resolution elements is then scanned. If the presence of a star is signalled in the second group within one second, the desired pattern recognition has thus been established. Thereupon the satellite changes its attitude such that the second guide star takes up a predetermined position *X*, enabling the astronomical observations to be carried out paraxially. The second guide star is tracked in the position *X* with the aid of the four adjacent resolution elements in the shape of a cross.

immediately after the yo-yo has done its work, and also for the 'discharge' of a reaction wheel. The output signals of the magnetometer can also be used for determining whether the satellite is 'stationary', i.e. whether it is no longer rotating about an axis aligned with the Sun.

The magnetometer consists of three orthogonal 'flux gates' or 'peaking strips' [6]. The magnetometer must not of course be disturbed by the magnetic fields of the three actuator coils. Compensation for these fields is provided by auxiliary coils.

A flux gate is a transformer with a strip core, whose primary coil is energized by a sinusoidal signal. The amplitude of this signal is such that the core just becomes saturated at the maximum energizing current, provided there is no external magnetic field. If, however, an external (weak) field appears in the direction of the core, the voltage across the secondary coil will be distorted. The amplitude of the second harmonic appearing in the secondary voltage as a result of this distortion is proportional to the external field. After filtering and phase-sensitive rectification, a d.c. voltage proportional to the field to be measured is therefore available.

The complete control system

Fig. 13 shows a block diagram of the complete automatic attitude-control loop. The onboard computer with its memory units (*OBC*, fig. 3) in conjunction with the attitude-control logic (*ACL, fig. 14*) acts as a comparator as well as a controller. Although these elements are inevitably somewhat complex in structure in view of the large number of logical decisions that have to be made for switching from one sensor to the other in the various operating modes, some simplification is nevertheless possible through the adoption of a

periodic sampling system, for which only one controller suffices. Given a sampling period of 1 second, the required accuracy of control can then still be reached. Of course the capacity of the computer must be large enough to ensure that the constant feed-in of all attitude data does not obstruct the processing of new setpoints and other instructions from the control room on the Earth as well as the processing of all the astronomical measuring data. As we saw in the preceding article [1], the satellite can only exchange these data with the ground station (via blocks *TC* and *TM*) once in each 12 hours — and the exchange has to be completed within ten minutes.

All that need be said here about the onboard computer is that it also provides the attitude-control system with clock-pulse signals by means of an 8 MHz oscillator with frequency-dividing stages. The computer memory unit comprises seven blocks, each of which can operate independently; one of them (which one is immaterial) is required for the storage of all the programs. The other blocks are available for data storage.

The attitude-control logic (*ACL*) comprises the magnetic-control logic (*MCL*), which controls the energizing of the three magnetic coils (*MC*) by making use of the output signal from the magnetometer (*MGM*) and the data from the reaction wheels. The coarse-control logic (*CCL*) can intervene in several ways; for example a signal from the coarse solar sensor (*CSS*) may lead to a demand for a torque to be supplied

[6] See H. Zijlstra, Experimental methods in magnetism, North-Holland Publ. Co., Amsterdam '1967, part 2, p. 37.
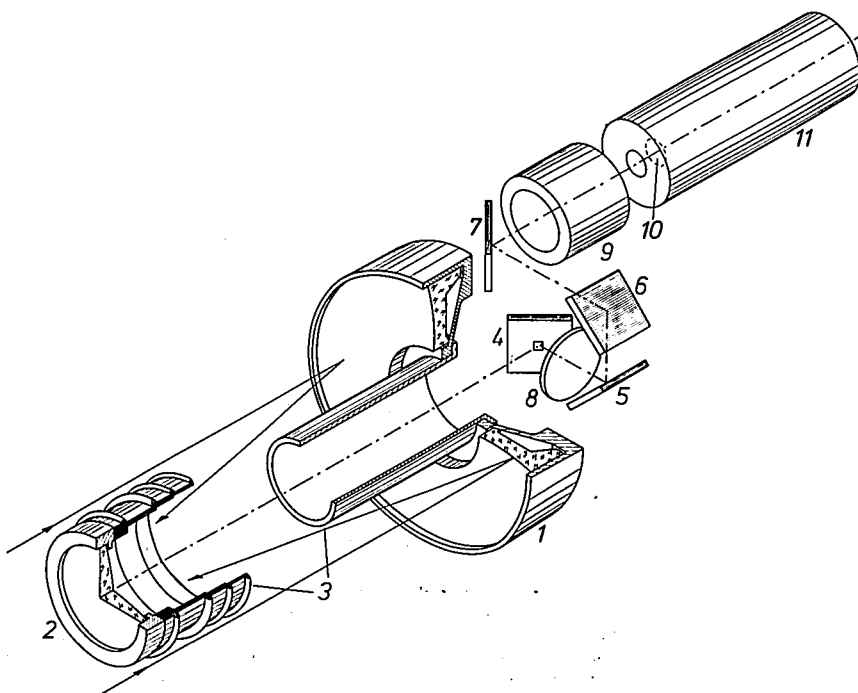
Fig. 12. The Cassegrain telescope, which forms part of one of the experiments on board the satellite. This telescope is also used for star tracking. *1* primary mirror. *2* secondary mirror. *3* light baffle. *4* oblique beam-folding mirror with central opening in the focal plane. *5-7* beam-folding mirrors. *8* field lens for enlarging the field of view. *9* relay lens. *10* photocathode. *11* image-dissector camera tube. The radiation entering parallel to the optical axis is the radiation on which the astronomical observations are made. This radiation reaches a spectroscope (not shown) through the small aperture in the mirror *4*. Radiation whose direction deviates slightly from that of the optical axis is transmitted to the photocathode of the star sensor by mirrors *4, 5, 6* and *7*. The star sensor generates a signal from which the onboard computer calculates the control signals for fine pointing and star tracking.
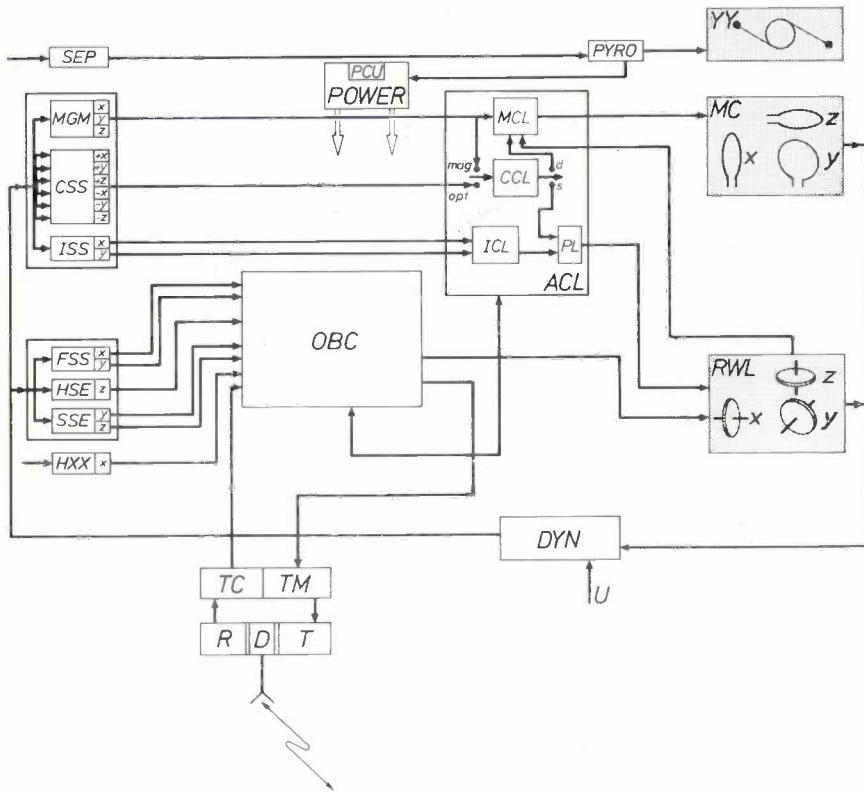
Sensors

MGM	magnetometer
CSS	coarse solar sensor
ISS	intermediate solar sensor
FSS	fine solar sensor
HSE	horizon sensor
SSE	star sensor
HXX	proportional counter for hard X-ray sources, functioning as sensor

Actuators

YY	yo-yo
MC	magnetic-coil system (torquer)
RWL	reaction wheel
SEP	time switch started on separation from launcher
PYRO	pyrotechnic actuator

Comparator and controller

OBC	onboard computer with storage
ACL	attitude-control logic
MCL	magnetic-control logic
CCL	coarse control logic
ICL	intermediate control logic
PL	priority logic

Connections

TC	telecommunications
TM	telemetry
R	receiver
T	transmitter
D	duplexer

Miscellaneous

POWER	power supply
PCU	power-control unit
DYN	satellite dynamics
U	external interference
d	despin
s	Sun acquisition
mag	magnetic
opt	optical
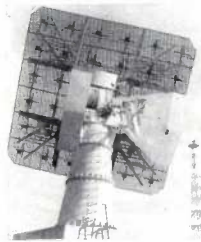
Photo ESRO

Fig. 13. Block diagram of the attitude-control system in the astronomical satellite ANS. The onboard computer and its memory units (*OBC*) and the attitude-control logic (*ACL*) together perform the tasks of comparator and controller; their operation can be modified from Earth via the ground station at Redu (Belgium). One of the ground-station aerials is shown in the diagram. The left-hand part of the diagram contains all the sensors, the right-hand part the three types of actuator.

via the magnetic-control logic (*MCL*). As can be seen in fig. 13, a signal from the intermediate solar sensor (*ISS*) produces an effect on the reaction wheels via its own control logic (*ICL*) and the priority logic (*PL*). The priority logic decides which of the two possible signals (from *CCL* and *ICL*) will control the reaction wheels. This decision depends on the operating mode of the satellite.

The operating modes

The various situations of the satellite after separation from the last launcher stage each have a corresponding operating mode, and these have to be completed in succession by the attitude-control system. It will be convenient to describe these modes in the

sequence in which they occur in a normal mission (fig. 1). In addition there are the 'eclipse' modes and a number of modes for dealing with emergency situations.

Table I gives a survey of the two groups, totalling ten operating modes. As noted earlier, the first task of the attitude-control system is to point the satellite at the Sun; it next has to carry out the fine pointing of the astronomical instruments.

Aiming at the Sun

The flight schedule, from the moment at which the satellite separates from the launcher — the 'mission profile' to which Table 1 corresponds, runs approximately as follows (see also fig. 13). After the initial spin has been reduced by the yo-yo the remaining
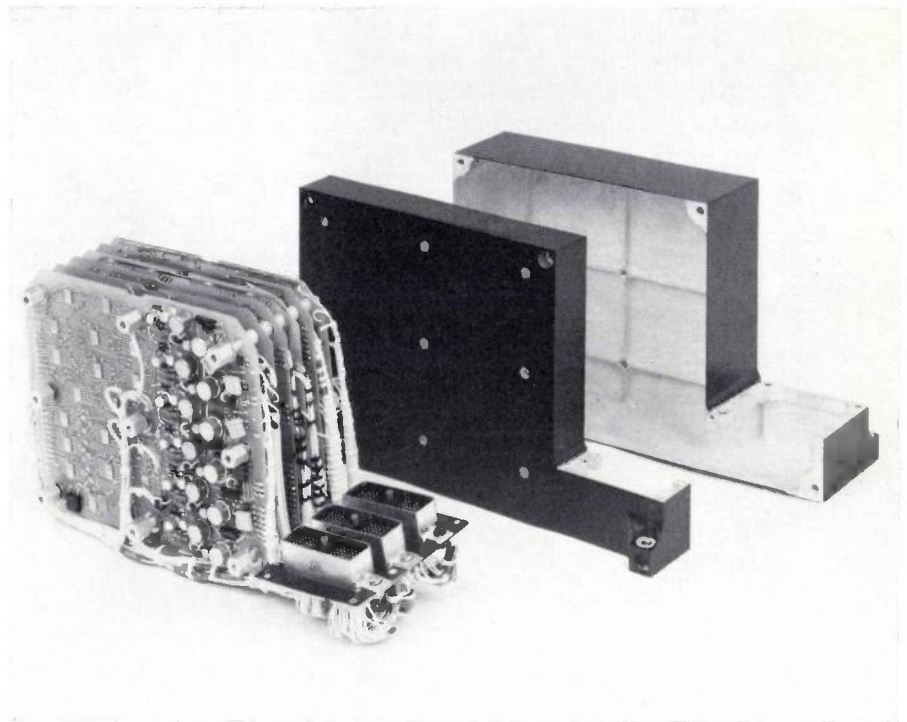
Fig. 14. The attitude-control logic circuitry. The nine relatively heavy screws are designed to help prevent vibration of the printed-circuit boards. They shift the natural vibrations of the stack towards higher frequencies, so that the loads encountered during the mission are relatively small.

Table I. The operating modes. In a 'normal' mission the attitude-control system goes through modes *1-6* after the satellite has separated from the launcher. The emergency modes *7* and *8* come in if the battery voltage is too low. Modes *9* and *10* come in when the satellite is temporarily in the Earth's shadow. The abbreviations used are defined in fig. 13.

Mode	Sensor	Actuator	Data storage	Task of magnetic control logic (*MCL*)	Setpoint speed of reaction wheels from
1 Despin	—	*YY*	yes	—	—
2 Solar detumbling	*CSS*	*MC*	yes	attitude control	—
3 Magnetic detumbling	*MGM*	*MC*	yes	attitude control	—
4 Sun acquisition	*CSS*	*RWL-x* *RWL-y* *RWL-z*	yes	discharging *RWL*	*ACL*
5 Sun pointing	*ISS* *MGM*	*RWL-x* *RWL-y* *RWL-z*	yes	discharging *RWL*	*ACL*
6 Fine pointing	*FSS* *SSE*	*RWL-x* *RWL-y* *RWL-z*	yes	discharging *RWL*	*OBC*
7 Emergency mode (during detumbling)	—	—	no	—	—
8 Emergency mode (during standstill)	*CSS* *ISS* *MGM*	*RWL-x* *RWL-y* *RWL-z*	no	discharging *RWL*	*ACL*
9 Eclipse (during detumbling)	*MGM*	*MC*	yes	attitude control	—
10 Eclipse (during standstill)	*CSS* *ISS*	—	yes	—	—

rotation is eliminated in the 'solar detumbling' mode by the magnetic coils (*MC*), under the control of the six coarse solar sensors (*CSS*). Even when the sensors no longer detect any change in lighting level, the satellite will still be able to rotate about an axis that makes a small angle with the connecting line between satellite and Sun. In the third · operating mode ('magnetic detumbling') the output signal of the magnetometer is therefore used instead of that of the solar sensor. If neither of these two sensors detects any further change during the next period of 50 seconds, the attitude-control logic concludes that the satellite is stationary, and the 'Sun-acquisition' mode can then begin. The result of this fourth mode is that the satellite points at the Sun with its positive z-axis within an angle of 30°. In this mode the reaction wheels are the actuators and the coarse solar sensors react to the value of the controlled quantity. At the end of this mode the solar panels are deployed. During the next mode ('Sun pointing') the z-axis is pointed at the Sun within an angle of 1°. The two intermediate solar sensors (*ISS* in fig. 13) supply the signal in this situation; the reaction wheels effect the change of attitude, backed up by the magnetic coils as soon as the speed of a wheel exceeds 20% of its maximum permissible value.

In the 'fine-pointing' mode, the last one in a successful mission, the positive z-axis will be pointed at the Sun and will remain pointed at it with an error of no more than 1 minute of arc — apart from a programmed misalignment provided by the offset facility. The two fine solar sensors (*FSS* in fig. 13) supply the signal. The next step in this mode is to point the astronomical instruments by means of a rotation about the z-axis. Since this mode is the one concerned in the actual scientific objectives of the project — the astronomical observations — a separate discussion on this is given below.

Modes 7 and 8 in Table I are designed to cope with emergency situations. The emergency measures are initiated by a signal that indicates that the battery voltage is too low. This signal can be generated by the power-supply units in the satellite and then passed on to the power-control unit (*PCU* in fig. 13). If the voltage is already too low during detumbling (modes 2 and 3) the controller breaks the connection between the battery and the satellite. The battery is then charged by the solar cells without having to supply current for the other equipment. This is possible because the solar panels have some chance of generating power during the tumbling of the satellite, even before they are deployed. An automatic command switches all the equipment on again when the voltage has increased to a high enough value.

If the same emergency situation arises when the satellite is stationary (mode 4) the same action is taken, but in this case the communication equipment (*TM* and *TC* in fig. 13) and the devices that keep the satellite roughly aimed at the Sun remain switched on.

Even if a fault arises in the onboard computer, the satellite is kept pointed at the Sun by the attitude-control logic (*ACL* in fig. 13). This does however increase the inaccuracy, since the attitude-control logic cannot work with the fine solar sensor (*FSS*) but only with the intermediate solar sensor (*ISS*). In the next contact between the satellite and the ground station an attempt can then be made to restore the original program. Of course, this will only help when the fault is in the software. If the hardware is also out of order, an adapted program may sometimes solve the problem.

If none of the six coarse solar sensors generates a signal, which means that ANS is in the Earth's shadow (modes 9 and 10), the attitude-control logic will deliver the eclipse signal, and power-saving measures will then come into operation. If this should already be the case during the satellite detumbling — which is only possible in the event of a serious orbit-injection error — the attitude-control logic then switches over to 'magnetic detumbling' (mode 3). This is the appropriate mode in this case, because the Sun is not required here as a reference or as a power source. After completion of the detumbling, all the equipment will be switched off except for the coarse and intermediate solar sensors; these sensors detect the end of the eclipse period.

Fine pointing of the x-axis

Once the z-axis has been accurately pointed at the Sun, fine pointing of the astronomical instruments at a celestial object is carried out in a number of operating modes that may be regarded as submodes of the fine-pointing mode (*Table II*). All possible transitions from one submode to another (*Table III*) take place on a command from — and under the complete control of — the onboard computer.

Pointing starts with the submode 'slew'. In this mode the satellite, controlled by the horizon sensor, turns from the previous direction of measurement to the new one. The object of this submode is thus to obtain a particular value for the angle between the x-axis of the satellite and the local vertical (see fig. 10*b*). In the next submode, 'scan', the angle obtained is kept constant, which implies that the satellite has to rotate about its centre of mass at the same angular velocity as that with which the centre of mass describes its orbit, which is 4 degrees per minute for an orbital period of 90 minutes. In this way the telescope, which looks along the x-axis, is able to scan a complete circle once in every revolution. Apart from the ordinary scan there is a slow-scanning submode in which the scan takes place

at the much lower angular velocity of 0.4 degrees per minute; during this mode the setpoint angle is continually updated under computer control.

When the star sensor recognizes the programmed pair of guide stars during one of the scanning modes, a signal to terminate this operating mode is sent to the onboard computer. Then fine pointing at the object under observation starts, so that it can be accurately tracked by the observation instruments; this finally brings the satellite into the situation in which the astronomical observations can be made. In this mode, referred to as 'star pointing', the onboard computer calculates the control signals for the reaction wheels,

Details of some control programs

In the 'slew' submode control is effected by electrical servomotors which drive the reaction wheels or, after reversal of the energizing current, apply a braking force to them. *Fig. 15* shows the block diagram of this minimum-time control system. As the name suggests, the aim of the control program is to bring about the desired slew in the shortest possible time. A control of this type is by nature nonlinear. A practical difficulty is that the perturbations are not easily predictable; friction is one of them, and another is that with increasing motor speed the torque delivered by a motor decreases as a result of magnetic saturation.

Table II. The control submodes for fine pointing (Table I, mode 6) in which the astronomical instruments are pointed at a stellar object by rotation about the z-axis. The positive z-axis remains continuously pointed towards the Sun. The abbreviations used are defined in fig. 13.

Submode	Rotation about x-axis y-axis z-axis controlled by			Angular velocity around z-axis	Inaccuracy
Slew	FSS	FSS	HSE	0-2 °/s	—
Scan	FSS	FSS	HSE	4 ± 2 '/s	0.3°
Slow scan	FSS	FSS	HSE	0.4 ± 2 '/s	0.3°
Star pointing	FSS	SSE	SSE	zero	1'
Offset pointing	FSS	SSE	SSE	zero, except during transition to new attitude	1'
X-ray pointing	FSS	FSS	HXX	zero	*

* Strongly dependent on strength of X-ray source; 1 minute of arc with strong sources.

using the signal from the star sensor as input. For hard X-ray sources it is also possible to use the output signal of the appropriate proportional counter as input signal for tracking. The satellite can enter this 'X-ray-pointing' mode from either the slow scan or the star-pointing mode. In the ordinary scan it will generally not be possible to 'catch' the X-ray source, whose intensity is usually too weak.

Normally the computer is programmed to point the observation instruments *at* the object of interest. It is however possible to switch from the star-pointing mode to an offset-pointing mode in which the instruments are deliberately pointed to one side of the object of interest to allow the background radiation to be measured.

The scanning submode is also the operating mode of the satellite when it enters into contact with the ground station. If for some reason the satellite cannot then be completely reprogrammed it is always possible to switch from this submode to an emergency program, for example to search for new X-ray sources elsewhere.

To conclude this section some details will be given of the control procedures used in the modes preceding fine pointing.

In such a situation the phase-plane method [7] is used for representing the relation between the angular error ε and its time derivative $\dot{\varepsilon}$. The angular error of the satellite is defined as the deviation from the desired attitude angle. In our case the phase plane (*fig. 16*) contains a number of trajectories which, when saturation and friction are left out of account, are parabolic in shape. In the figure the driving branches are shown blue and the braking branches red. These are the only trajectories along which the system can move.

Table III. The permitted transitions between the operating submodes mentioned in Table II; the transitions are made at a command from the onboard computer.

Slew
Scan
Slow scan
Star pointing
Offset pointing
X-ray pointing
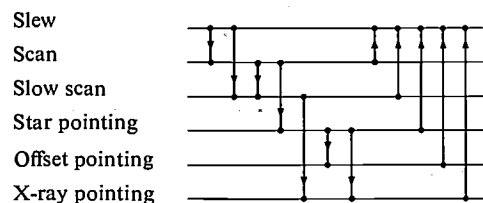
[7] See for example: J.-C. Gille, M. J. Pélegrin and P. Decaulne, Feedback control systems, McGraw-Hill, New York 1959, chap. 25.

If the computer is only switched over from driving to braking at an angular error $\varepsilon = 0$, a system with an initial error $-\varepsilon_i$ will describe an oscillatory movement along the associated parabola l_i and $l_i{}^*$. A limit cycle of this kind is of course not permissible because the angular error has a permanent maximum value ε_i.

The result required can be achieved by making the computer switch over earlier, i.e. before the angular error ε has been reduced to zero. With a single switching operation the system can then be moved into the 'ideal' braking branch (l_p in fig. 16). This takes the system straight to the origin of the phase plane; in the programming for the onboard computer the trajectory l_p is represented by the expression $\alpha\varepsilon + \dot\varepsilon|\dot\varepsilon| = 0$. The branches thus satisfy the equations $\varepsilon = \pm (\dot\varepsilon)^2/\alpha$, where α is a gain factor.

An added complication is the fact that the angular velocity $\dot\varepsilon$ must in all circumstances be limited to a value at which rotation of the satellite due to the reaction wheel does not cause any interfering signal in the control circuit of one of the two other wheels. The onboard computer counteracts such gyroscopic effects by preventing the reaction wheel from generating a control torque as soon as $\dot\varepsilon$ exceeds a specified value. Consequently the system can never move outside the area bounded by the two dashed lines in fig. 16. It will therefore not reach the ideal braking branch at point P but at Q, via the dashed line. The presence of a trajectory of this kind, on which the control torque is zero, also makes the problem of saturation much less serious.

In practice of course certain imperfections are encountered; for example, near the origin the curve is exceptionally steep because the maximum control torque is operative in that region. This could give rise to severe overshoot and unstable stationary behaviour. In this stage the control torque should therefore be reduced to keep the remaining limit cycle — which is unavoidable with an on-off control — within acceptable bounds. The magnitude of the torque is chosen in the second block of the diagram (fig. 15), the maximum value per wheel being 0.015 Nm. The rest of the program is such that from the point where the absolute value of the angular error is 3° and its derivative is approximately 0° per second, the control torque decreases in direct proportion to the angular error. The ideal braking branch in the phase plane must also be adjusted accordingly, and the computer does this by appropriately changing the gain α (fig. 15). The system can then move sufficiently close to the origin and remain there. The slew mode ends as soon as the system reaches a point where the angular error is 0.8° and its derivative 0° per second.

In the scanning mode the scanning speed — which of
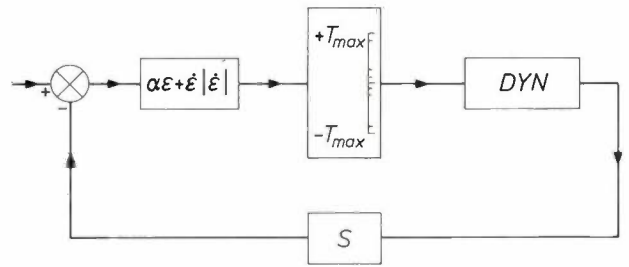
Fig. 15. Block diagram of the minimum-time control of a reaction wheel. The onboard computer switches an electric servomotor from *driving* to *braking* action on the reaction wheel as soon as the relation between the angular error of the satellite and its time derivative can be expressed by the equation $\alpha\varepsilon + \dot\varepsilon\dot\varepsilon = 0$, which is held in the program; ε is the angular error of the satellite with respect to the desired attitude angle, and α is a gain factor (depending on the control torque). *DYN* satellite dynamics. *S* sensor. The magnitude of the control torque T is chosen in the second block (extreme value T_{max}) in dependence upon the angular error.
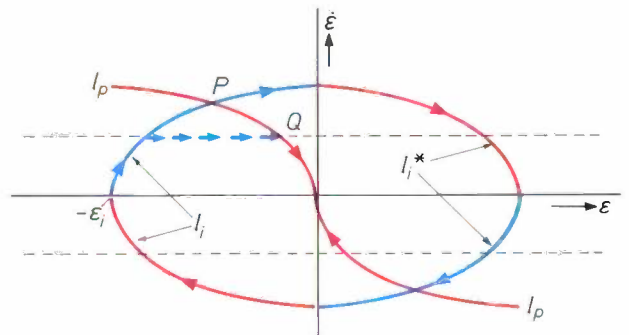
Fig. 16. Phase-plane diagram of the control system in fig. 15. When magnetic saturation and dynamic frictional behaviour are neglected, families of *parabolic* trajectories lie in the phase plane (driving branches are blue, braking branches red). The parabolae l_i, $l_i{}^*$ form the limit cycle that occurs when the switchover from driving to braking is made only upon the initial error $-\varepsilon_i$ being reduced to zero; this situation is not permissible. Earlier switchover (at position P) makes it possible for the system to return directly to the origin ($\varepsilon = 0$, $\dot\varepsilon = 0$) via the ideal braking branch l_p. To avoid undesirable gyroscopic coupling with the controls of the two other reaction wheels, the value of the angular velocity $\dot\varepsilon$ must always remain inside the region between the two dashed lines. Since the computer does not allow the servomotor to generate any further control torque as soon as this limit is reached, the system will in practice return to the ideal braking branch not at P but at Q, via the dashed line.

course must be equal to the speed of the satellite in its orbit around the Earth — must be exactly 4 minutes of arc per second. The degree of constancy required is so high that the angular measurement once per second with the horizon sensor (*HSE*) does not provide sufficiently accurate information. This difficulty is circumvented by including in the calculation the seven preceding angular values: the computer determines the new measured value with the aid of a curve which is fitted as closely as possible to the eight ($= 7 + 1$) values. After subtracting the setpoint value from this new measured value a deviation is obtained which is then subjected to a PID control algorithm [8]. The final

Fig. 17. A test bench (*TB*), suspended virtually without friction, is used for investigating parts (*P*) of the attitude-control system. This arrangement simulates as closely as possible the physical conditions which the satellite will encounter during the mission. *W* wire for torsionless suspension (consists of 20 insulated strips). *SC* slip-ring contacts for electrical connections. *K* terminals. *TOS* torsion sensor. *DIF* difference amplifier which converts the measured torsion into a control signal. *SER* servomotor that makes the suspension point of the wire rotate with the bench (so that the torsion angle remains below 0.1°). *B* shield tube. *D* displaceable weights for adjusting the moment of inertia of the bench plus test arrangement. *AB* air bearing; this prevents oscillation of the bench about horizontal axes. *S* securing points, which protect the bench in the event of a sudden collapse of pressure in the air bearing.

accuracy achieved is 0.1° for the angular deviation and 0.02 °/s for its derivative.

The slow-scan mode requires even more complicated algorithms, for one reason because variations in the angular velocity with which the satellite travels in its orbit around the Earth give rise to control difficulties in this mode. 'Slow' implies that the scan takes place at a speed which is an order of magnitude smaller than the angular velocity of the satellite in its orbit (Table II). This is achieved by periodically adapting the setpoint value of the attitude angle. Because of these frequent stepwise adjustments the control cannot have an integrating action — otherwise, for example, excessive overshoot could occur. The attitude of the satellite may then continue to show a fixed deviation with respect to the 'sliding' setpoint value, but within certain limits this is not relevant to the quality of the measurements to be performed. The attitude data required for evaluating the astronomical results are regularly measured with the desired accuracy and stored in the memory units.

Test procedures

The control algorithms chosen and the operation of all equipment are tested in two ways, first with software tests and secondly with hardware tests. The software tests consist of computer simulations — first with one and then with three axes — in which the attitude control and the dynamics of the satellite are simulated in a single program. A more detailed discussion of this is however out of place here.

The hardware tests are carried out on all the components of the equipment. They are performed with the equipment rotating about one axis only, since the torques with which the system works are so small that a rotation sufficiently free from friction can only be obtained about a single axis.
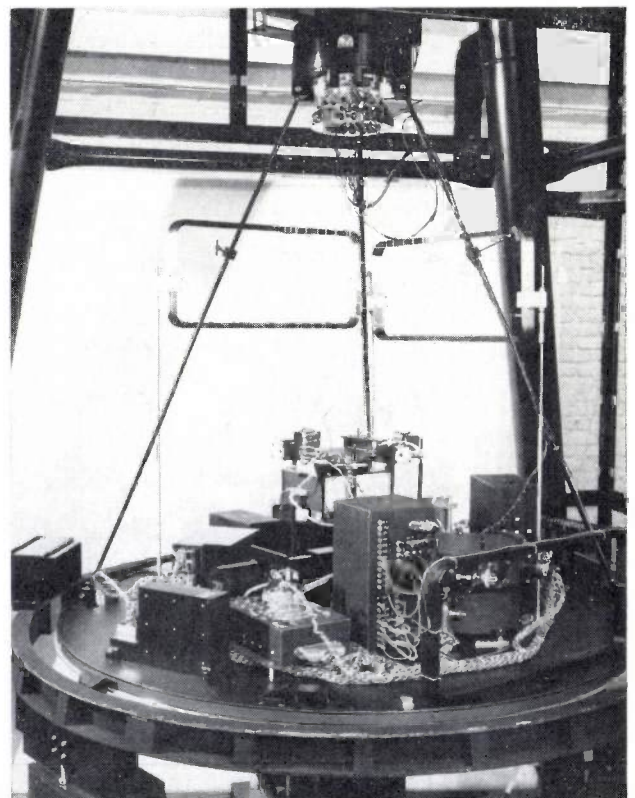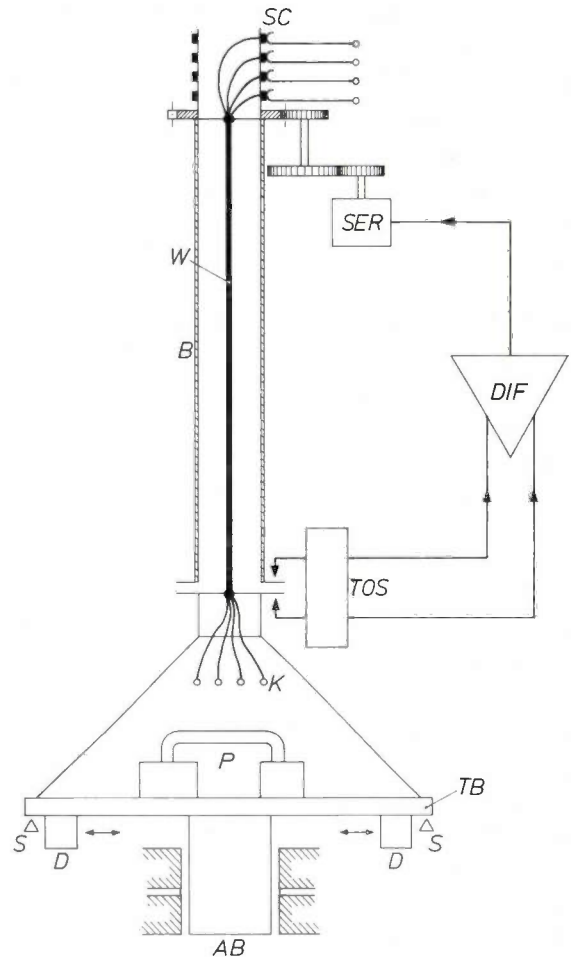
[8] The controller has a proportional (*P*), an integrating (*I*) and a differentiating (*D*) action.

In these hardware tests the conditions which the satellite will encounter during its mission are simulated as closely as possible. Frictionless rotation is simulated as well as the electromagnetic radiation for the sensors: visible radiation for the Sun sensors and infrared for the horizon sensor.

A collimated beam of high luminous intensity is used to simulate the Sun, two beams of low luminous intensity simulate the two guide stars, and there is a horizon simulator. Three pairs of Helmholtz coils are used to simulate the Earth's magnetic field for every attitude of the satellite. This arrangement contains the magnetometer and the three coils of the magnetic actuator.

Frictionless operation is simulated by means of a test bench suspended by a wire two metres long, which has a very low torsion constant (*fig. 17*). The components to be investigated are placed on the bench, and care is taken to ensure that the moment of inertia of the bench and components is equal to that of the satellite about the same axis.

To reduce the torsion constant of the suspension wire the wire consists of twenty very thin strips of beryllium copper, which slide against each other on twisting. The strips are insulated by a plastic coating, so that they can be used for supplying power to the components on the bench and for the data exchange. To prevent the torsion from reaching an undesirably high value, the suspension point of the wire is rotated by a servomotor together with the bench. The maximum twisting angle of the suspension wire is thus kept below 0.1°. Other factors, such as air friction and the slight imbalance of the bench give an 'interfering' torque, which has to be added to the torsion of the wire. The total applied torque is somewhat smaller than the torque that will act on the satellite in orbit. The interfering torque can be varied as required by means of an accurately calibrated coil on the bench. This enables realistic conditions to be simulated in an easily controllable way.

Since the satellite has three principal axes of inertia with different moments of inertia, it was necessary to make the moment of inertia of the bench adjustable as well. This is done by means of sliding weights under the bench, enabling the moment of inertia to be varied from 5.5 to 17.5 kg m².

Oscillations about horizontal axes are avoided by means of an air bearing under the bench; the bearing carries no load and has a very high stiffness (approximately 6×10^7 N/m).

Since the maximum permissible control-loop error in the ANS system is 1 minute of arc, the attitude of the bench has to be determined with much greater accuracy. This is done by means of a measuring system with optoelectronic readout, which was developed for numerically controlled machine tools [9]. The resolution of this system is 1 μm, which in our case corresponds to an angle of 1 second of arc.

Summary. The Netherlands astronomical satellite (ANS) is equipped with a digital attitude-control system (sampling period 1 s), controlled by an onboard computer (storage capacity 28 000 words of 16 bits) which also stores measurement programs and results. The computer software can be modified from Earth. The attitude-control system starts by pointing one axis of the satellite at the Sun; this is the axis perpendicular to the solar panels and the optical axis of the astronomical instruments. The satellite can rotate about this axis through any desired angle and can track celestial objects to an accuracy better than 1 minute of arc. There are three solar sensors (coarse, intermediate and fine), a horizon sensor (sensitive to infrared), and a star sensor. The star sensor 'recognizes' and tracks a programmed pattern of guide stars, seen through the telescope of one of the astronomical experiments on board, since the stars to be observed are often too faint to serve as a reference. The actuators used are three orthogonal reaction wheels and a set of magnetic coils that cause the satellite to rotate by means of the Earth's magnetic field. A 'yo-yo' actuator eliminates the initial spin of the satellite. The operating modes for a normal mission, the emergency modes and the simulation of flight conditions in test procedures are discussed.

[9] The linear measuring system LMS III of the Philips Industrial Equipment Division, Eindhoven. This system was developed for measuring distances in numerically controlled machine tools.
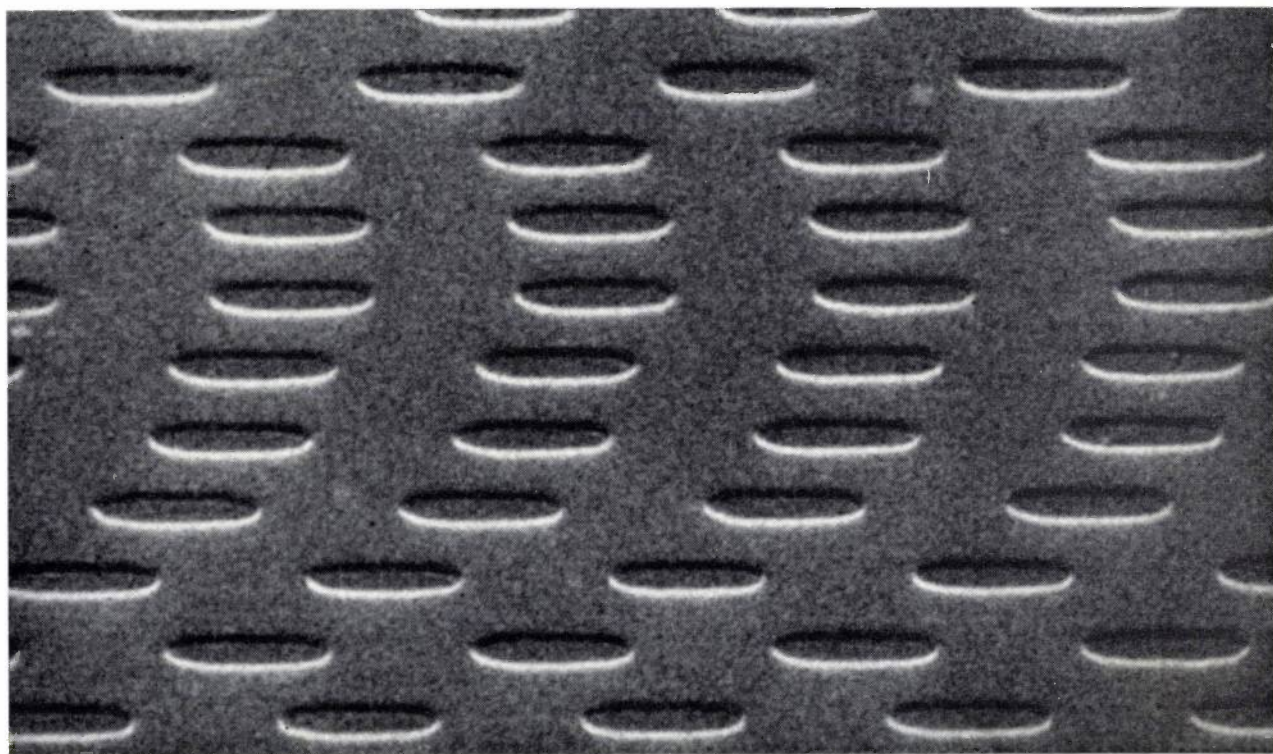
Several years ago Messrs Compaan, Forsthuber and Burgstede of the Education Group of the Philips Electro-Acoustics Division (ELA) in Eindhoven put considerable thought into the idea of using some form of 'gramophone record' as an inexpensive and therefore rapid way of disseminating visual information; a project that to some extent had been prompted by Mr Wols. One of the schemes put forward in the Education Group was taken up by a team of specialists from Philips Research Laboratories; this team was later expanded to include members from several other Product Divisions.

The result of these joint efforts is the Philips 'VLP' (video long-play) system, which was first shown to the international press at a conference held at the Research Laboratories on 5th September 1972. 'VLP' records can be pressed in much the same way as normal gramophone records. A colour-television programme lasting for more than thirty minutes can be recorded on a single disc.

The first of the four articles that follow explains how the information is recorded on a Philips video long-play record and how it is scanned by the playing equipment. The other three articles are about signal processing during both recording and playback, the optical system used in the player, and the control systems. These articles cannot of course give much more than a picture of the current state of development. It also seems too early to undertake a detailed comparison with other systems that have been proposed.

The commercial reproduction of documents and the commercial reproduction of recorded sound are both activities whose later progress was almost beyond the imagination of the early pioneers. We shall therefore attempt no forecast of the future of the commercial reproduction of recorded moving pictures, but will merely say that the Philips 'VLP' system will meet a clearly established requirement.

178

Philips tech. Rev. 33, 178-180, 1973, No. 7

The Philips 'VLP' system

K. Compaan and P. Kramer

Now that almost every home and many educational institutions have a television set it is natural to think of the possibility of using it, in combination with a play-back unit, for reproducing programmes that have been permanently recorded in some way or another. This gives the user the freedom of being able to watch a programme he is interested in at a time convenient to himself — the same freedom he can enjoy with a shelf of books or a collection of gramophone records.

The 'VLP' system [*] described here allows a colour-television programme of about 30 minutes duration to be reproduced from a recording on a 'gramophone record' 30 cm in diameter, the usual size for a long-playing record. The 'VLP' record can be produced simply and in quantity by the normal pressing techniques. The 'VLP' system is complementary to the video cassette recorder (VCR), which has been on the market for some time, but to some extent it offers an alternative to it. A programme can be recorded as desired with a cassette recorder, but it is more expensive to produce recorded tapes than it is to press 'VLP' records.

The development of the 'VLP' system is the result of the combined efforts of a team of specialists in very divergent fields. In this article we shall give a broad general survey of the system; the three short articles that follow will describe some of the components in more detail [1] [2] [3].

The information is recorded on the record disc along a spiral track, which occupies the part of the disc between the 10 cm and 30 cm diameters. The speed at which the disc rotates has been made equal to the picture frequency, 25 s⁻¹ for the European market and 30 s⁻¹ for North America. As we shall see later, this offers some interesting possibilities. If the playing time is half an hour, these figures give a pitch of 2 μm for the track.

For following a track with such a small pitch an optical method is very suitable. In the 'VLP' player this scanning is done with a spot of light 1-2 μm in diameter, projected on to the track by a lens.

The diameter of the spot is of the same order of magnitude as the wavelength of the light used in the equipment, and it is therefore no longer possible to speak of a particular diameter. A diffraction pattern (an Airy disc) is formed at the focal plane of

Drs K. Compaan is with the Philips Electro-Acoustics Division (ELA), Eindhoven; Dr P. Kramer is with Philips Research Laboratories, Eindhoven.

the lens; this pattern consists of a central maximum surrounded by successive dark and light rings. To produce a pattern in which the half-intensity diameter is 0.9 to 1.0 μm at the wavelength used, a lens with a numerical aperture of 0.4 is required.

The information for the reproduction of a television picture is recorded as a succession of short grooves or pits of variable length and repetition frequency. The width of the pits is 0.8 μm, and the depth 0.16 μm (see the title photograph). Since in pressing a gramophone record the surface roughness does not amount to more than 0.01 μm, it is clearly a practical possibility to make such a pattern in the surface of a pressed disc.

If the spot of light falls on the surface of the disc between two of the pits, then most of the light will be reflected back into the objective lens. If on the other hand the spot falls on one of the pits, the light will be deflected by diffraction at the pit in such a way that most of it is not returned to the objective (fig. 1). In this way the intensity of the light reflected through the aperture of the lens is modulated by the pattern of pits [1]. The intensity variations are converted into an electrical signal by a photodiode. The width and depth of the pits in the surface are arranged to give as large a modulation depth as possible.

To obtain a high signal-to-noise ratio in the detector signal, the reflected beam should have as high an intensity as possible. If the photocurrent is too low, the noise will no longer be mainly determined by the thermal noise in the detector, but by the shot noise in the photon current. We have therefore used an He-Ne laser as the light source. Also, to improve the reflectivity, the surface of the 'VLP' disc has been coated with a thin layer of evaporated metal.

Some of the members of our team have developed a special technology that enables the He-Ne laser to be manufactured in quantity. This 1 mW laser has been built into the player in such a way that it can be of no possible danger to the user.

The information on the surface of the disc can be read out through a transparent protective layer. Any contamination or damage only affects the outer surface of this layer, and not the disc. The diameter of the beam at this outer surface is much larger than the spot, so that these imperfections have very little effect on the detector signal. This arrangement makes use of the

very small depth of focus of an objective lens with a resolving power in the micron range.

To enable it to be encoded in the pattern of pits, the video signal undergoes a number of special processes [2]. The bandwidths of the brightness signal and the colour signal are both limited to some extent. The frequency of the colour-signal subcarrier, which is 4.43 MHz in the PAL system, is reduced to a value of 1 MHz, fixed with respect to the line frequency. This allows the original carrier frequency to be restored when the record is played, even if there are deviations caused by variations in the speed of revolution. The sound is treated as a frequency modulation of a 250 kHz carrier. The brightness signal, which modulates a 4.75 MHz carrier, determines the *repetition frequency* and the *average length* of the pits, while the preprocessed colour and sound signals give a *modulation of the length* of the pits.

Work has also been done on other encoding systems whose potentialities include the recording of a video signal with a wider bandwidth.

The master record from which the moulds are produced for pressing the 'VLP' records is cut by a laser in the specially prepared surface of a glass disc. This cutting is done at the same speed at which the records
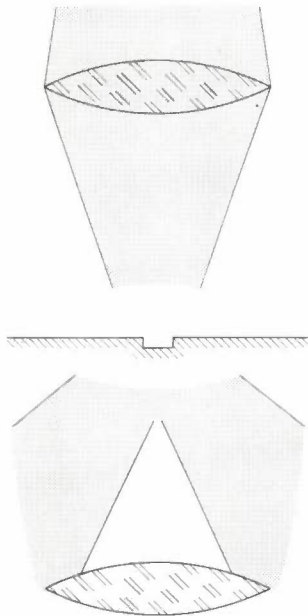
Fig. 1. Modulation of the light by a pit in the surface of a 'VLP' record. For clarity the system is drawn as if the record were transparent, with the beam incident from above and a second lens placed underneath the record to receive the light. The pit is also shown many times enlarged with respect to the rest of the figure. If the record surface is flat, all of the incident light is received by the lower lens. If there is a pit in the surface there will be diffraction, and some of the light will be deflected; when the pit is correctly dimensioned much of the incident light will be deflected away from the aperture of the lower lens. In practice the record surface is reflecting, and only one lens is required for concentrating the light on to the record and receiving the reflected light.

will be played. A scene can therefore be recorded on the record directly from the video camera or transferred without delay from a magnetic tape. The moulds are made in the usual way from the master by an electroplating process.

If a 'VLP' player is to give good results four special requirements have to be satisfied. In the first place, the speed of revolution of the record must be kept constant to an accuracy of 1 in 10^3, or the playback of the video signal will be unsatisfactory.

Secondly, the objective must remain focused on the surface of the record. Because of its large aperture the objective has only a very small depth of focus. Although the irregularities on the surface of the record are locally very small, the deviations over a wider area can be as much as 0.5 mm.

In the third place the beam of light must remain centred on the track, even though the track may be not truly circular (out-of-round) or eccentric. Deformation of the disc during pressing can lead to out-of-roundness; eccentricity of the spindle-hole in the record and play between it and the shaft of the playback unit can cause the track to rotate eccentrically. The player must be able to operate correctly even when the total deviation of the track from the ideal position is as much as 0.1 mm.

Finally, the complete optical system must move radially across the record at the rate at which the track advances ('tracking'), without the aid of a continuous groove or other mechanical guide in the disc or the player. To meet these requirements a number of control systems have been developed; these will be described in one of the following articles [3].

Fig. 2 shows a diagram of the 'VLP' player. The complete pick-up unit can move backwards and forwards on a carriage on rails underneath the record disc *1* to follow the track. The light from the laser *2* is focused at the record by the objective *3*. The control systems mentioned above act on the objective and a pivoting mirror *4*, thus keeping the beam focused and centred on the track. A prism *5* ensures that light reflected by the record falls on the detector *6*.

The 'VLP' player can also be used to show the pictures in reverse motion, slow motion or at higher speed. This is possible because the record rotates synchronously with the picture frequency — 25 rps for the European version, 30 rps for the American one. Consequently at each rotation of the track the field-synchronizing pulses always fall within two fixed diametrically opposite sections of the record disc.

[*] Trade Mark, property of N.V. Philips' Gloeilampenfabrieken.
[1] G. Bouwhuis and P. Burgstede, this issue, p. 186.
[2] W. van den Bussche, A. H. Hoogendijk and J. H. Wessels, this issue, p. 181.
[3] P. J. M. Janssen and P. E. Day, this issue, p. 190.

(A television picture consists of two interlaced fields.) Wherever the spiral track crosses the two sectors it therefore contains the same information — the field-synchronizing signal. This means that inside the sector the beam can be allowed to change from one turn of the

turns. This will give a 'picture-book' of about 45 000 different pictures. Address coding allows any particular picture to be found rapidly.

The large number of pictures — which can be completely different if desired — that can be stored on the
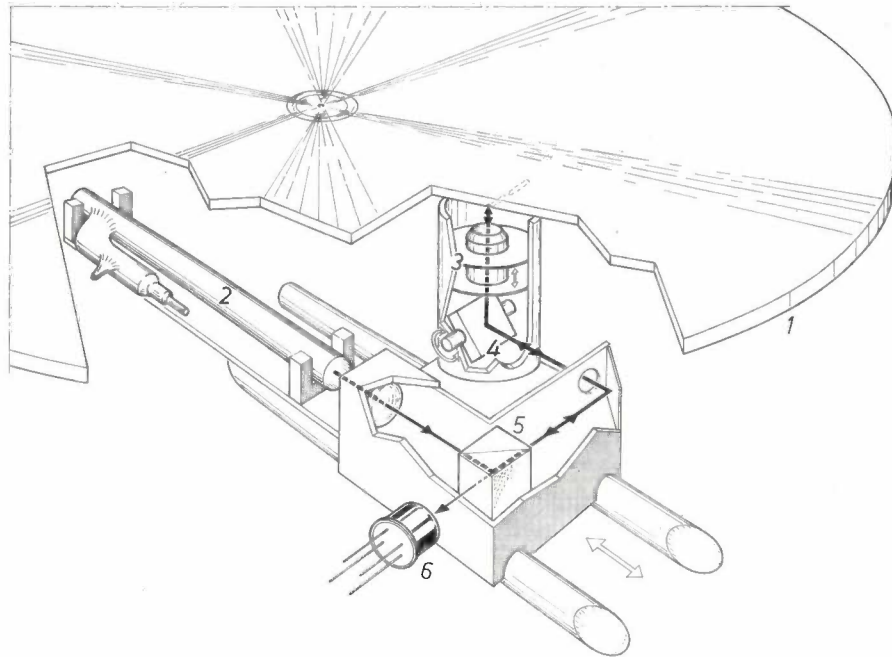
Fig. 2. Schematic diagram of the 'VLP' playback unit. The record *1* is scanned from below by light from the He-Ne laser *2*. The objective *3* is held focused on the record by a system based on a loudspeaker mechanism. The pivoting mirror *4* ensures that the beam remains centred on the track; the mirror is operated by a rotating-coil arrangement. Incident and reflected light are separated by the prism *5*. The detector *6* converts the reflected light into an electrical signal.

track to an adjacent one, without spoiling the picture. This is done by applying a control pulse at the correct moment to the control system for correct centring on the track. By continually repeating the same turn and thus the same picture in this way, a stationary picture will be obtained. By repeating each picture twice a picture in slower motion will be obtained, and by omitting every other picture the action of the scene will be reproduced at twice the speed. A picture in reverse motion is obtained by jumping back a turn at each half revolution.

Because of the accurate centring of the scanning beam on the track the cross-talk between successive turns is very small (< -30 dB), so that it is possible to record completely different pictures on successive

'VLP' record, and the scope for manipulation of the recorded information, make the 'VLP' system one that clearly offers more than the simple dissemination of video information.

Summary Television pictures are recorded on the Philips video long-playing ('VLP') record in a spiral track of pits in the surface. The pits have constant width and depth but the lengths and spacings are variable. The information is read out by a beam of light, which is reflected at the surface of the record. The reflected beam is modulated by deflection of the light through diffraction at the pits. To enable the 'VLP' playback unit to operate at the required accuracy, control systems have been developed for holding the speed of rotation of the record constant, focusing the read-out beam on the record surface and centring the beam on the spiral track without the assistance of mechanical guides. The player can be used to show the recorded pictures one at a time, and will also allow them to be shown in reverse motion, slow motion, or at faster speed.

Signal processing in the Philips 'VLP' system

W. van den Bussche, A. H. Hoogendijk and J. H. Wessels

On the 'VLP' record there is a single information track in which all the information is stored for the reproduction of a colour-television programme with sound. The photographic process used for writing the information on to the master record is highly non-linear, and can therefore only be used for recording a signal with just two levels. The information can then only be present in variations in the distances between successive transitions from one level to the other. A system of coding will therefore be necessary if four or five signals are to be recorded simultaneously in this way (brightness, two colour signals and one or two sound signals).

As is customary in television techniques, the two colour signals are processed as a single signal: a quadrature-modulated subcarrier, as in the NTSC and PAL systems [1]. At playback a reference signal is required for decoding; this must have exactly the same phase as the original subcarrier. Special measures have therefore been taken to ensure that the 'VLP' record player provides a reference signal that meets the specified requirements.

In the coding process the brightness signal, which modulates the frequency of a carrier, is the principal signal. The colour and sound signals, each modulating a carrier of lower frequency than the principal signal, are contained in symmetrical displacements of the zeros of this principal signal. The fully coded signal is recorded on the record as a pattern of short grooves or pits of the same width and depth, but variable length and spacing. On average the total length of the pits occupies exactly half of the total track length. This feature is made use of in the control system [2] that guides the read-out beam along the track.

We shall now describe the systems for recording and playback, and explain the operation of the electronic units with the aid of block diagrams.

Recording

At a speed of 25 rps and with our coding system, the central part of the record will not reproduce correctly the part of the normal video spectrum (fig. 1) lying above 3 MHz. The original colour information present in this part of the spectrum thus disappears completely. For the parts of the track that are further out from the centre of the record this cut-off frequency has a higher value, so that colour information can be directly recorded and reproduced there. However, this directly recorded colour information would cause an annoying interference with the colour information, which as will be seen later, is recorded in such a way as to enable corrections to be made on playback for frequency variations in the colour carrier caused by variations in the speed of rotation of the record. By using filters to cut off the brightness spectrum at 3 MHz before recording, annoying interference can be avoided. In
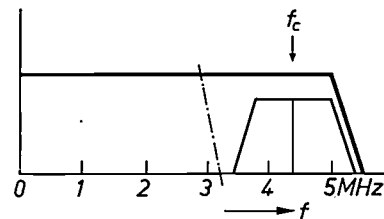
Fig. 1. The video spectrum in the PAL colour-television system. The brightness signal occupies the frequency band from 0-5.5 MHz. This band also contains the colour information, which modulates a carrier of frequency f_c (4.43 MHz). In the 'VLP' system the video spectrum is cut off above 3 MHz, while the colour carrier is shifted to 1 MHz.

addition, the record can now be used with a minimum track diameter of 10 cm, which gives a considerable increase in playing time.

The brightness signal, limited in bandwidth, modulates the frequency of a 4.75 MHz carrier; the frequency swing of this modulation is made less than the bandwidth of the information to be transferred. This gives sidebands that are wider than the frequency sweep. However, the higher-order sidebands of the high-frequency signals can be neglected, because of their small modulation index. It is therefore sufficient to record a frequency band extending 3 MHz on either side of the carrier. This means that there is bandwidth available below 1.75 MHz for colour and sound signals.

The frequency of the colour subcarrier is therefore reduced from 4.43 MHz to 1 MHz, in the case of the

Ir W. van den Bussche, A. H. Hoogendijk and Ir J. H. Wessels are with the Philips Audio Division, Eindhoven.

[1] F. W. de Vrijer, Philips tech. Rev. 27, 33, 1966.
[2] P. J. M. Janssen and P. E. Day, this issue, p. 190.

PAL system, while the total bandwidth of the encoded colour signal is limited to $\pm$ 0.5 MHz. The sound information modulates the frequency of a 250 kHz carrier, with a frequency sweep of 75 kHz. If desired a second sound channel can be included at the low-frequency end of the spectrum for stereo purposes or for a spoken text in another language.

The three signals processed as described above for brightness, colour and sound are combined in an amplitude ratio of 10 : 2 : 1. The amplitudes of the signals for colour and sound can be smaller since the noise in these signals is less because of the smaller bandwidth. (Apart from spectral-distribution effects, the noise in a particular band is proportional to the bandwidth.)

In the resulting signal, whose frequency spectrum is shown in *fig. 2*, colour and sound can be considered as artificial lower sidebands in a single-sideband modulation of the brightness signal as the carrier. Symmetrical amplitude limiting of such a single-sideband-modulated signal results in the synthesis of the missing upper sidebands at the expense of the power in the lower sidebands. The amplitude ratio therefore becomes 20 : 2 : 1 after limiting.

Looking at it in another way, the symmetrical limiting gives a signal of rectangular pulses in which the brightness information is contained as frequency modulation, while colour and sound information give a symmetrical modulation of the width of the pulses (duty-cycle modulation, *fig. 3*). This rectangular-pulse signal is suitable for recording in a pattern of pits on the 'VLP' record.

It now only remains to say something about the way in which a stable frequency is obtained for the colour subcarrier in spite of speed variations that may occur when the record is played. The frequency used for the colour subcarrier on the record is 1 MHz. This is exactly 64 times the line frequency in the video signal. By locking the two frequencies together, the colour information can be recreated at playback on a stable 4.43 MHz carrier obtained from the line frequency and the frequency of a stable 4.43 MHz oscillator.

We shall now discuss, with the aid of a block diagram, the electronic equipment used in recording.

Block diagram

Fig. 4 gives a block diagram of the electronic circuit used in recording a 'VLP' record. After what we have said above, no further explanation is required of the processing of the sound signal by the audio amplifier *1* and an FM modulator *2*.

The part of the brightness signal to be recorded is separated out from the video signal by a lowpass

filter *3*. The FM modulator *4* is a multivibrator driven by the modulation; it gives a rectangular-pulse signal modulated in frequency. However, if the duty-cycle modulation described earlier and shown in fig. 3 is to function properly, then the modulated brightness signal must not have too short a rise time. To ensure this the lowpass filter *5* is included, which makes the signal almost sinusoidal.

The filter *6* separates the colour information from the video signal. The variable-gain amplifier *7* ensures a constant level for the colour signal, with the reference level derived from the 'bursts', constant-amplitude signals consisting of a number of oscillations at the colour-carrier frequency, which follow each line-synchronizing pulse. These bursts are added to a colour-television signal as a phase reference for decod-
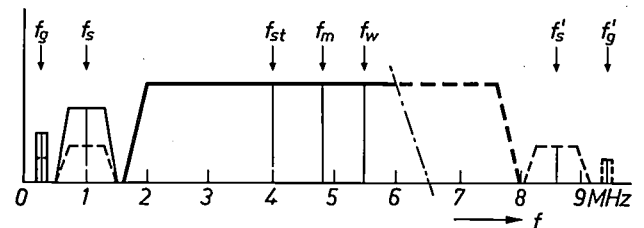
Fig. 2. The frequency spectrum recorded on a 'VLP' record. Solid lines: the spectrum after summing the brightness, colour and sound information; dashed lines: the spectrum after symmetrical limiting. The relative amplitude ratios of the various components are not shown to scale. f_{m} centre frequency of the brightness-signal spectrum. f_{w} frequency corresponding to the white parts of the picture. f_{st} frequency corresponding to the peaks of the synchronizing pulses. $f_{\mathrm{s}}, f_{\mathrm{s}}'$ the 1 MHz colour subcarrier and the corresponding frequency arising in the upper sideband on symmetrical limiting. $f_{\mathrm{g}}, f_{\mathrm{g}}'$ the 250 kHz sound carrier and its counterpart in the upper sideband.
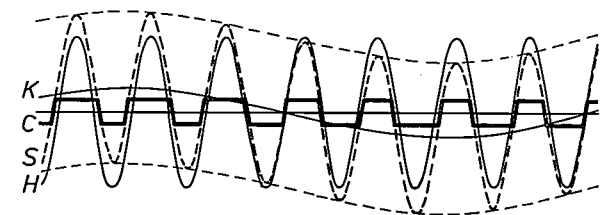
Fig. 3. Schematic diagram of the waveforms in the 'VLP' system. *H* brightness signal, *K* colour and sound signal; the three signals each modulate the frequency of a carrier. *S* superposition of brightness, colour and sound signals. *C* the rectangular-pulse signal produced on symmetrical limiting; the brightness information is still present here as a frequency modulation. Because of the finite rise time of *H* the colour and sound signals give a duty-cycle modulation of the rectangular-pulse signal.

ing the colour signal. The bursts are observed by the burst detector *11* at a command from the synchronization separator *10*. If there is no colour signal and hence no bursts, the gain is reduced to zero, so that interfering signals in the colour circuit cannot cause coloured spots in the picture.

The phase discriminator *12* synchronizes the 4.43 MHz crystal oscillator *13* with the subcarrier frequency given by the bursts. This combination forms the 'electronic flywheel' normally used in television receivers to make sure that the subcarrier is available with the correct frequency and phase between bursts. A second oscillator *17* provides the 64th harmonic of the line frequency. This oscillator is locked to the line frequency by dividing the frequency of its output by 64 (*16*), and comparing the result with the line frequency (*15*).

The mixer *14* gives the sum frequency of the two oscillators, 5.43 MHz. This sum-frequency signal is mixed in the mixer stage *8* with the colour signal on the 4.43 MHz carrier, after which the lowpass filter *9* separates out the difference-frequency component.

Playback

The output signal from the photodetector of the 'VLP' record player must be translated back again so that it can be played back via a monitor or television set. Apart from demodulating the brightness and sound signals the most important operations here are the transformation and correction of the subcarrier for the colour signal, and correcting the 'drop-outs' on the record as well as possible — 'drop-outs' are places where the signal is missing because of damage to the record. For playback on a television receiver the signal also has to modulate a suitable UHF or VHF carrier so that the aerial socket can be used as the input. We shall now look at the electronic circuit of the record player with the aid of the block diagram shown in *fig. 5*.
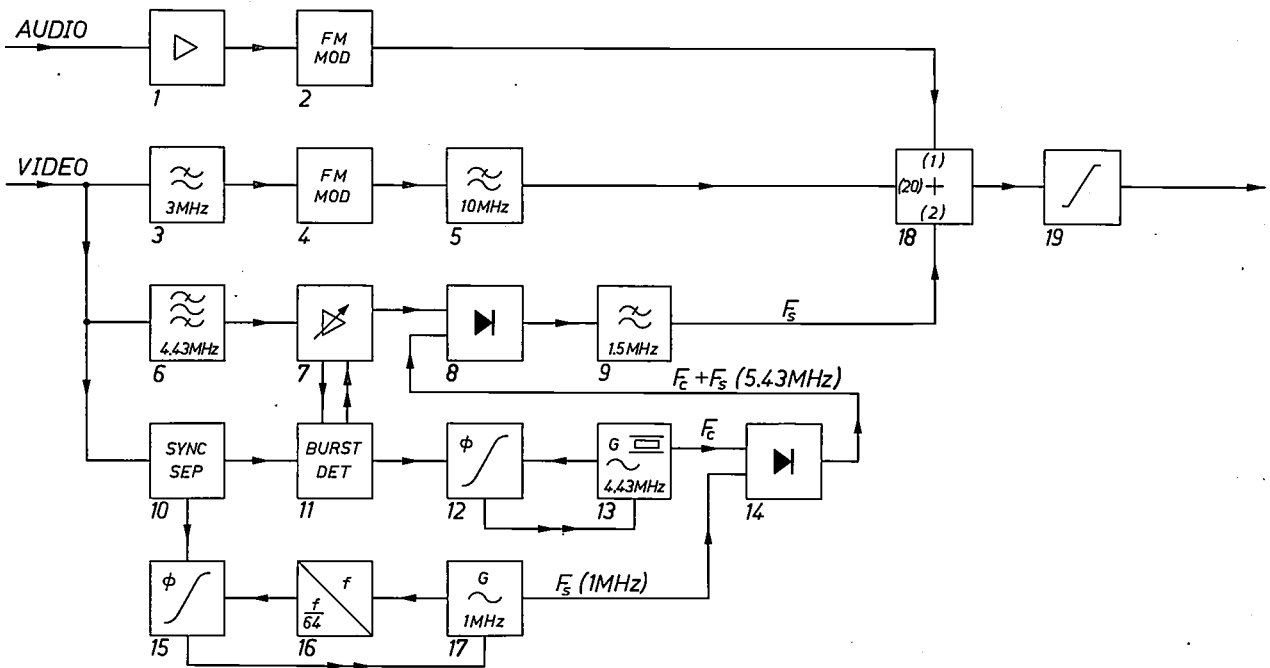
Fig. 4. Block diagram of the electronic circuit used in recording a 'VLP' record.

1 Audio amplifier	*8* Mixer stage	*15* Phase discriminator
2 FM modulator	*9* Lowpass filter	*16* Divider circuit
3 Lowpass filter	*10* Synchronization separator	*17* Oscillator
4 FM modulator	*11* Burst detector	*18* Summation circuit (the ratios in which the components are added is shown)
5 Lowpass filter	*12* Phase discriminator	
6 Bandpass filter	*13* Crystal oscillator	
7 Variable-gain amplifier	*14* Mixer stage	*19* Symmetrical limiter

This is the colour information, modulating a 1 MHz carrier.

After the three signals have been added in the correct ratios (*18*), they go through the symmetrical limiter *19*. The signal is then suitable for supply to the light modulator used for recording on the 'VLP' record.

Sound, colour and brightness signals are separated from each other by filters (*2* to *7*). The brightness signal (frequency between 1.5 and 6 MHz) then goes through the correction amplifier *8*, which gives a falling linear characteristic. In the symmetrical limiter, which follows next, a purely frequency-modulated signal is produced.

Here signals are generated at the high-frequency end of the spectrum at the expense of the signals at the low-frequency end, which were emphasized in the preceding amplification on account of this process. As well as this main channel, the brightness signal also goes through an identical subsidiary channel (*11, 12, 13*), which gives a signal delayed with respect to the main channel by exactly the time of one line scan. The

signal is produced whose frequency is exactly half that of the frequency-modulated brightness signal. The drop-out detector *18* reacts to this and the switch *14* is operated during a period of 3 μs, so that the brightness of the preceding line is employed instead. This 3 μs is found to be a sufficiently long time in practice since very few drop-outs of longer duration are encountered. The transit time of the signal through the
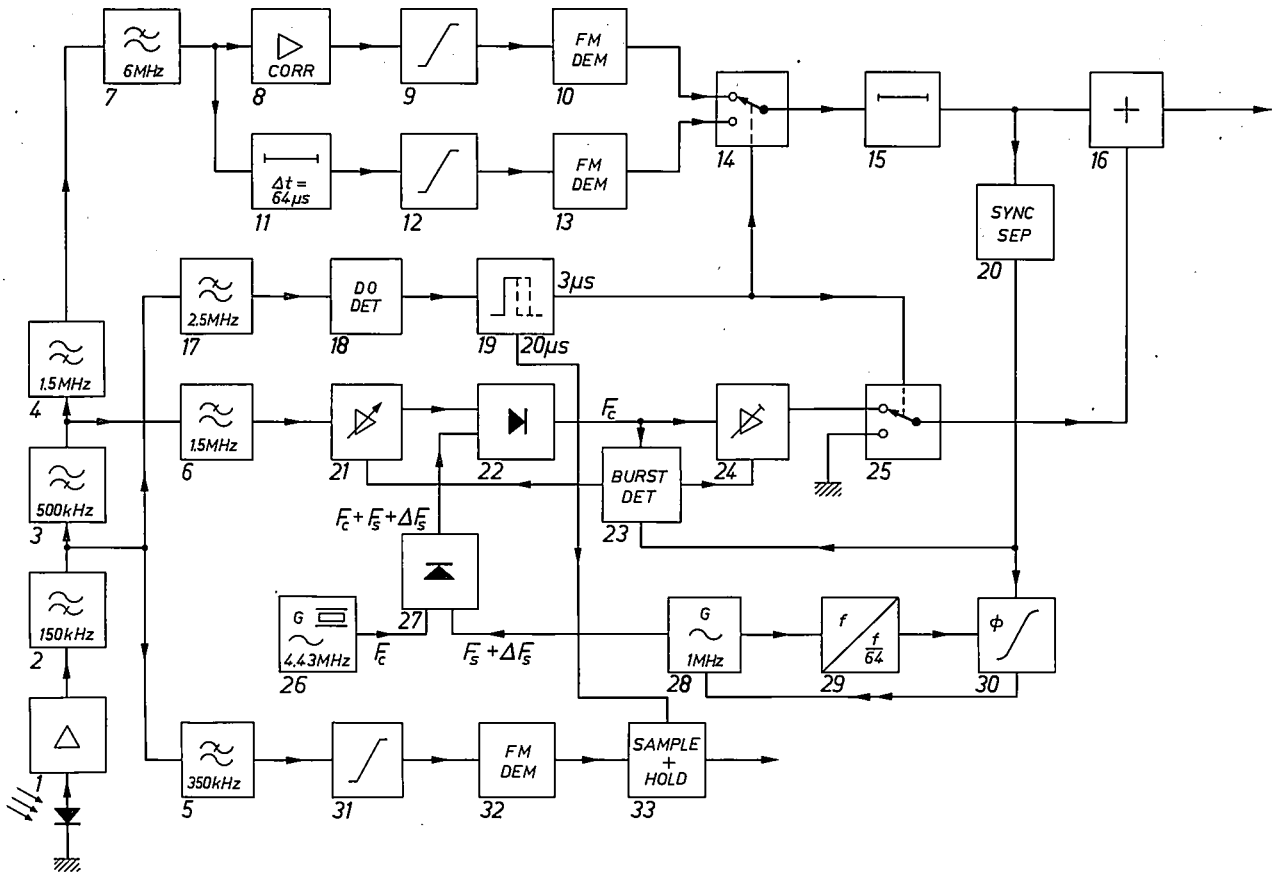
Fig. 5. Block diagram of the electronic circuit of the 'VLP' playback unit.

1 Photodetector and preamplifier
2-7 Filters
8 Correction amplifier
9 Symmetrical limiter
10 FM demodulator
11 Delay line
12 Symmetrical limiter
13 FM demodulator
14 Electronic switch
15 Delay line
16 Summation circuit
17 Lowpass filter
18 'Drop-out' detector
19 Monostable circuit
20 Synchronization separator
21 Variable-gain amplifier
22 Mixer stage
23 'Burst' detector
24 Switching amplifier
25 Electronic switch
26 Crystal oscillator
27 Mixer stage
28 Oscillator
29 Frequency-divider circuit
30 Phase discriminator
31 Symmetrical limiter
32 FM demodulator
33 Correction circuit for 'drop-out'

transmission characteristic of the delay line is corrected in such a way that a falling linear characteristic is again produced. In this way if a drop-out should be encountered the signal from the previous picture line is available as a replacement signal.

The filter *17* separates out the part of the signal for which the frequency is lower than 2.5 MHz. This is used to detect drop-outs. If the detector misses a pit a

demodulation circuit is longer than the time required by the drop-out detector, so that the switch *14* is already changed over before the drop-out in the signal arrives there.

The most complicated processing is that undergone by the colour signal. The line amplifier *21* gives a constant level for the signal that has been separated out by the filters *3* and *6* (frequency between 0.5 and

1.5 MHz). As in recording, the blocks *28*, *29* and *30* again form an electronic flywheel, now driven by the line frequency of the playback signal, which reproduces the 1 MHz carrier, including the frequency deviations that have arisen during playback. The free-running crystal oscillator *26* gives a 4.43 MHz signal that will now have to function as the new colour carrier. The mixer stage *27* gives a 5.43 MHz signal, containing any frequency errors that may be present. The difference between this frequency and the frequency of the colour signal reproduced from the record is formed in the mixer stage *22*. Since frequency deviations of the two signals cancel out, the desired stable 4.43 MHz carrier is recreated.

The burst detector *23* provides the reference signal for the variable-gain amplifier *21* and at the same time ensures that in the absence of a colour signal the colour channel is inoperative ('colour killer', *24*). If there is a drop-out the switch *25* short-circuits the colour circuit. Because of the averaging with the signal of the preceding line in PAL receivers, the missing colour fragments are now filled in at half saturation. 'Spikes' are thus prevented.

The sound signal is separated out by the filters *2* and *5* (frequency between 150 and 350 kHz) and is demodulated (*32*) after symmetrical amplitude limiting (*31*). This signal then also passes through the circuit *33*, which counteracts spikes in the signal level at drop-out.

Summing the colour and brightness signals gives the complete video signal. A delay line *15* is used here to correct for differences in transit time in the two circuits. As was stated earlier, audio and video signals then have to be used to modulate a carrier for playback on a normal colour television receiver.

Summary. To be able to record all the video and audio information in a single track on the 'VLP' record a number of signal processings are required, particularly since the recording has to be done with only two signal levels. The brightness signal, whose bandwidth is limited to 3 MHz, is used to modulate a 4.75 MHz carrier, the carrier of the quadrature-modulated colour signal is shifted to 1 MHz and the sound information is supplied as frequency modulation of a 250 kHz carrier. Summation of the signals processed in this way in the ratios 10 : 2 : 1 and symmetrical limiting of the sum signal give a rectangular-pulse signal. This contains the brightness signal as a pulse-frequency modulation and the colour and sound as 'duty-cycle' modulation. The new colour-carrier frequency of 1 MHz is exactly 64 times the line frequency. By linking these two frequencies together a stable 4.43 MHz colour carrier can be restored at playback, independently of variations in the speed of rotation of the record.

The optical scanning system of the Philips 'VLP' record player

G. Bouwhuis and P. Burgstede

The main function of the optical scanning system in the Philips 'VLP' record player [1] is to read out the video information recorded on the surface of the disc. To follow the information track on the record with the aid of a control mechanism it must also deliver a signal that is a measure of the lateral misalignment of the optical system in relation to the track. The video information is stored on the surface of the record in the form of small pits which are 0.8 μm wide, about 0.16 μm deep and of variable length. The pits are inscribed on a spiral track with a pitch of approximately 2 μm. The picture quality and mechanical accuracy of an optical system that can detect each pit individually and track the spiral sufficiently accurately to prevent cross-talk between adjacent turns of the track must meet extremely exacting standards. At the same time the system must be as simple as possible to produce and easy to adjust. Finally, care must be taken to ensure that the signal-to-noise ratio is sufficiently high. The article describes the optical system which we chose for the 'VLP' record player to satisfy these various requirements.

Read-out of information

To read out the information, a beam of light is focused on the track by means of a lens with a numerical aperture of 0.4. The diameter of the spot is approximately equal to the theoretical minimum for a lens with this aperture. This is because the aberrations of the lens chosen are negligibly small, so that the intensity distribution in the spot and the dimensions of the spot are determined not by the laws of geometrical optics but solely by diffraction at the lens aperture. The light source is a laser, which means that the intensity distribution over the entrance pupil of the lens varies in a radial direction. This together with the diffraction effect at the lens aperture results in an intensity distribution over the spot such that at the wavelength of the laser light (0.63 μm) the half-intensity diameter (the 'diameter' of the spot) is about 0.9 μm.

The light is reflected by the surface of the record, picked up by the lens again and concentrated on to a photodiode. Diffraction occurs at the pits in the sur-

face, and most of the diffracted rays fall outside the lens aperture (fig. 1). Less light is consequently received when a pit passes in front of the lens than when a smooth section of record surface does so. The pits thus modulate the current through the photodiode.

The spot diameter of 0.9 μm, due to diffraction at the lens aperture is, as we have just seen, approximately equal to the theoretical minimum value obtainable with this lens. Conversely, a detail of this size on the surface of the record would give diffracted rays that would only just fall within the lens aperture. Finer details give diffracted rays of greater divergence, which will therefore not fall within the lens aperture. This is in fact the case with the pits in the record surface since they are narrower than the diameter of the spot. Rays diffracted by the pits must consequently fall outside the lens aperture.

For maximum modulation of the detector current by the pattern of pits on the record the incident light that is reflected from the bottom of a pit must have a 180° phase difference from the light that is reflected from the record surface around the pit, and the intensity of both must be the same. The 180° phase difference is obtained by making the depth of the pits a quarter of the wavelength of the light used. This depth does not require great accuracy, however; the intensity reflected into the lens as a result of a small deviation from the correct value is a square-law function of the
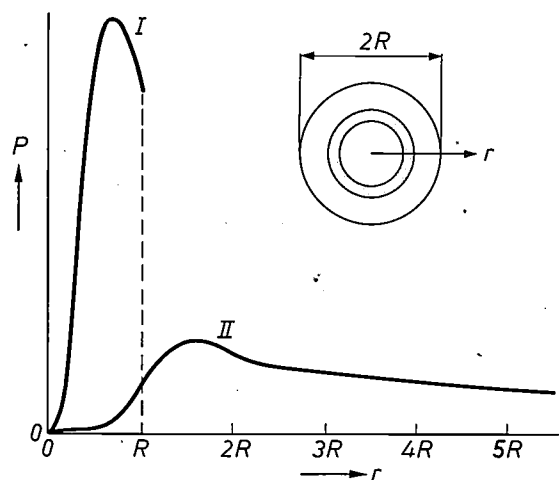
Fig. 1. The power P reflected by the surface of the record into the pupil of the objective plotted as a function of the distance r from the optical axis. Each curve shows the total power in a thin annulus of radius r. The diameter of the pupil is $2R$. Curve I represents the power for reflection from a flat part of the record, curve II the power when the spot is centred on a pit in a track in which the period of the pattern of pits is 1.5 μm and the width of the pits 0.8 μm; the pitch of the track is 2 μm.

G. Bouwhuis is with Philips Research Laboratories, Eindhoven, and P. Burgstede with the Philips Electro-Acoustics Division (ELA), Eindhoven.

deviation (*fig. 2*). Equality between the intensities is achieved by dimensioning the spot and the pits in such a way that the same amount of light falls outside a pit as falls inside it. In this way an optimum relationship is ensured between the track geometry and the intensity

Fig. 2. A slight deviation δ in the 180° phase difference between the light (amplitude A) reflected from the bottom of a pit and light reflected at the record surface beside it gives an out-of-phase component with an amplitude A_u. Now $A_u = A \sin \delta \approx A\delta$. The corresponding intensity is therefore $A^2\delta^2$. Since δ is directly proportional to variations from the nominal value $\frac{1}{4}\lambda$ for the depth of the pit, the remaining intensity is proportional to the square of the deviation from the correct depth.

distribution over the spot. In other words, for every track geometry there is an optimum value for the lens aperture used. The most important factor in choosing the aperture, however, is the highest frequency to be read from the record by the optical system.

Numerical calculations [2] show that with a decreasing period Λ of the pattern of pits along the track the reflected intensity decreases more rapidly than the modulation depth M of the detector signal (*fig. 3*). The definition of modulation depth used here is $M = (I_{max} - I_{min})/(I_{max} + I_{min})$. Modulation measurements on a prototype of the record player showed the same qualitative behaviour of intensity and modulation depth as a function of Λ as that found by calculation for an idealized situation.

For a track with a diameter of 10 cm (the innermost track on the record) at a frequency of 7 MHz, corresponding to $\Lambda = 1.12$ μm, a modulation depth of 55% was calculated. The highest modulation depth so far measured on the innermost track at 7 MHz is 15%. The measured and calculated values differ because a number of conditions were idealized in the calculations (e.g. the focusing and centring of the beam on the track were assumed perfect while the pits were assumed to present a perfectly rectangular profile along the track).

The cross-talk between adjacent tracks can be found by measurement. It can however also be calculated by adding together the contributions from the various regions (*fig. 4*) to the detector signal. In the calculation the track on which the beam is centred is assumed to have a different 'pit frequency' from the tracks on either side. The components with the two corresponding frequencies can then be calculated from the signal by Fourier analysis. A cross-talk level of −50 dB was measured for tracks with a width of 1 μm, combined

with the light beam with a half-intensity diameter of 1 μm mentioned above; this measured level gives good agreement with the calculations.

Until now we have always assumed that the spot was centred in relation to the track. If it is not, the modulation depth will decrease and the absolute level of the cross-talk signal will increase. Numerical calculations show that deviations of the spot from the track of up to about 0.2 μm will not give any noticeable cross-talk. This estimate has been confirmed by measurements.

The light source

The only existing light source suitable for our purposes was the He-Ne laser with its high brightness $(3.10^5 \text{ W/cm}^2 \text{ sr})$ and low noise at MHz frequencies. With this laser and a metallized record surface to increase reflection, it is possible to obtain a signal level such that the noise from the light source, the photodiode and its amplifiers give no unsurmountable problems. What is necessary is to ensure that not too much of the light modulated by reflection from the record surface finds its way back into the laser, because this kind of feedback would cause undesirable fluctua-
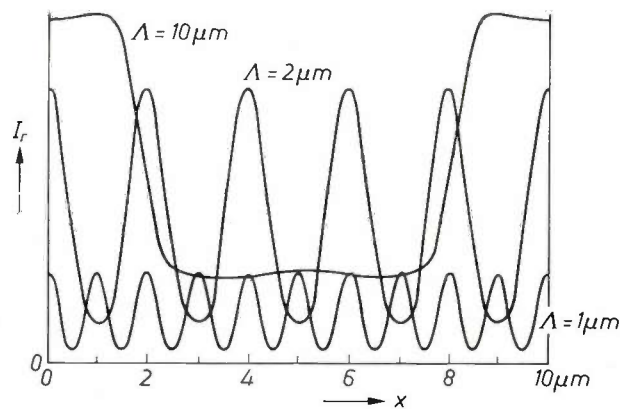
Fig. 3. Calculated modulation of the intensity I_r of the reflected light for various values of the period Λ of the pit pattern as a function of the position x on the track. $x = 0$ corresponds in all cases to a point half-way between two pits. For the purpose of the calculation the aperture of the lens used is taken as 0.4, the width of the pits as 0.8 μm and the wavelength of the light as 0.63 μm.
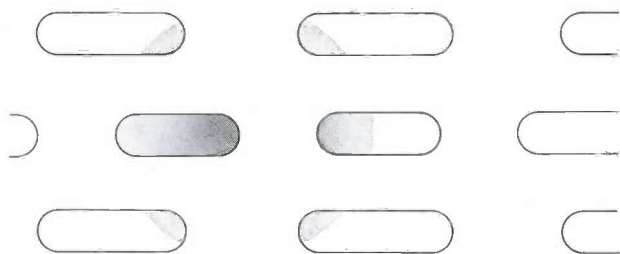
Fig. 4. Cross-talk between adjacent tracks. The density of the hatching gives some idea of the weighting factors which have to be applied in calculating the contributions to the reflection from the various pits.

[1] A general description of the record player is given in the article by K. Compaan and P. Kramer, this issue, p. 178.

[2] This and subsequent numerical calculations are largely based on a mathematical analysis by Prof. H. H. Hopkins of the University of Reading, England.

tions in the power output of the laser. The large distance over which laser light remains coherent also necessitates caution. Stray light reflected by moving parts can interfere with the primary beam at the detector and modulate the detector current very strongly.

As we have already noted, the intensity distribution over the spot on the record is determined not only by the lens aperture but also by the intensity distribution over the entrance pupil of the lens. The Gaussian distribution describing the radial change of intensity in the laser beam can be made wider or narrower at will by introducing an intermediate lens. While it is true that by doing so and using only a part of the laser beam we have less power available, the spot on the record is reduced in size owing to the more homogeneous distribution of the light over the entrance pupil. This enables the system to detect higher frequencies in the information on the record and thus improves the bandwidth of the 'VLP' system as a whole. It is thus a question of finding the most suitable compromise.

We have matched the laser to the entrance pupil of the objective in such a way that approximately 80% of the laser power is utilized. The remaining inhomogeneity of the light distribution results in a bandwidth which is $7\frac{1}{2}\%$ smaller than would be obtained with completely homogeneous illumination of the pupil. This bandwidth is however 16% larger than that which would be obtained if only 1% of the laser power fell outside the pupil.

The accuracy with which the distance from the objective to the record is kept constant for focusing purposes was investigated by means of numerical calculations [3]. As *fig. 5* shows, it is desirable that deviations from the optimum distance should not be greater than about 2 μm.
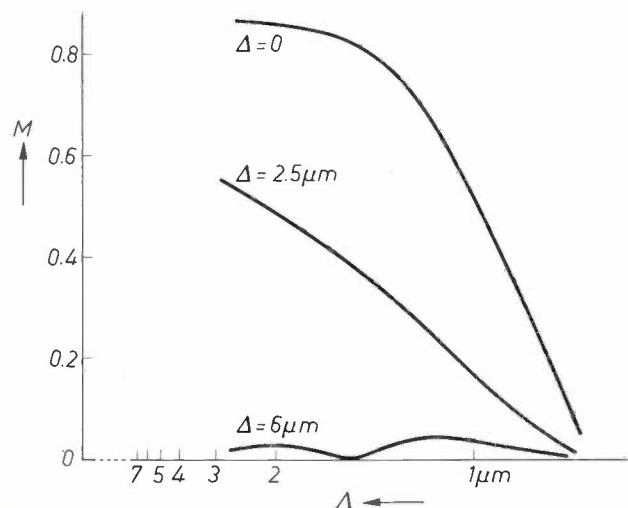
Fig. 5. Numerically calculated behaviour of the modulation depth M of the light reflected from the track as a function of the period Λ of the pit pattern for various deviations Δ from optimum focusing ($\Delta = 0$).

Tracking control signal

Since the scanning beam has to remain centred on the track to an accuracy of 0.2 μm, a requirement which cannot be satisfied with even the most accurate purely mechanical guidance system, a tracking-control system has been incorporated in the record player [4]. The information on the track can only be read out optically and the deviation of the beam from the centre of the track can therefore also only be measured optically. For this we use two auxiliary beams of light, each of which is imaged on its own detector by the optical system after reflection from the surface of the record.

The two auxiliary beams strike the track at equal distances in front of the main spot and behind it. Both beams are slightly displaced from the centre-line of the track, in opposite directions, so that each is partly on and partly alongside the track (*fig. 6*). The average current through the detectors of the two auxiliary beams depends on the deviation of the beams from the centre of the track. The difference between the two detector signals is applied to a lowpass filter with a cut-off frequency of 20 kHz and the output signal forms a suitable error signal for the control system.
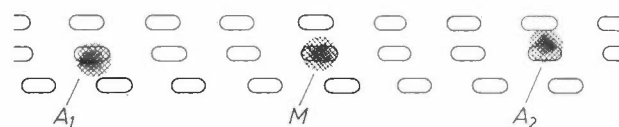
Fig. 6. Position of the main beam M and the two auxiliary beams A_1 and A_2 in relation to the track to be scanned.

The auxiliary beams are obtained by splitting the laser light into three approximately equally intense beams by means of a diffraction grating. (This process also gives rise to a number of weak unwanted beams which are removed; the available power is used with an efficiency of about 75‰.) The fact that the three beams are obtained by diffraction and not, for example, with mirrors, ensures that both the intensity relationships and the spacings remain constant.

Finally, in *fig. 7*, we present a schematic diagram of the complete optical system. As already mentioned, an intermediate lens (L) ensures that the laser beam is imaged on the entrance pupil of the objective in the desired fashion. It also produces a beam path such that the spot on the record is imaged on the detector without any other aids. This means that the detector can easily be adjusted and the sensitivity of the arrangement to small displacements or vibrations is reduced. A further point is that the useful radiation falls on a small part of the detector, thereby reducing the risk of modulation by stray light as mentioned earlier. Generally speaking, stray light will not be focused in the

same way as useful light, and only where the useful beam and the stray light coincide will interference occur. Interference results in intensity variations due to the phase variations of the stray light caused by reflection from moving parts. The stray light that does not coincide with the useful radiation merely creates a background which is constant and therefore much less objectionable.

A transverse magnetic field in the laser ensures that the laser beam is linearly polarized, parallel to the field [5]. This means that with a quarter-wave plate and a polarizing mirror [6] the incident and the reflected light can be effectively separated with the result, firstly, that the laser power can be used efficiently and, secondly, that undesired feedback of the modulated light to the laser is prevented.

[3] A theoretical treatment of a similar situation, the imaging of a black-and-white object instead of the light modulation by a phase object dealt with here, will be found in: M. Born and E. Wolf, Principles of optics, 3rd edition, Pergamon, London 1965.

[4] A description of the control system used in the 'VLP' record player is given in the article by P. J. M. Janssen and P. E. Day, this issue, p. 190.

[5] H. de Lang, Physica 33, 163, 1967.

[6] See for example: H. de Lang and G. Bouwhuis, Philips tech. Rev. 30, 160, 1969.

Fig. 7. Diagram of the optical scanning system. R 'VLP' record. B laser beam. G grating for producing the auxiliary beams. L lens for adapting the laser beam to the entrance pupil of the objective Obj. The intermediate image given by lens L also makes it readily possible to focus the light reflected from the record on to the detector Det. The polarizing mirror P, in conjunction with the quarter-wave plate W, separates incident light from reflected light. M pivoting mirror for following track on record. Inset a shows the positions of the three spots in relation to a track on the record, while inset b gives an impression of the three detectors with the three beams reflected from the record focused on them.

Summary. An optical system is used to read out the information which is stored in the form of a pattern of small pits in the surface of a 'VLP' record. The intensity of a beam of light reflected from the surface of the record is modulated by diffraction at the pits. The information is contained in the variation of the length of the pits and the distance between them. Their width determines the optimum value of the aperture of the lens used to focus the beam at the record and also to gather the reflected light. A numerical aperture of 0.4 requires a width of 0.8 μm. The depth of the pits is a quarter wavelength. The relationship between the modulation depth of the reflected light and the period of the pattern of pits is found from numerical calculations. Similar calculations also yield the permissible tolerances on the centring and focusing of the beam on the track to be scanned. The only light source of sufficient brightness is a He-Ne laser. Two auxiliary beams of light obtained from the main beam with a diffraction grating enable the system to supply a reference signal for the control system that keeps the scanning beam centred on the track.

Control mechanisms in the Philips 'VLP' record player

P. J. M. Janssen and P. E. Day

The playing time and the bandwidth of the recorded video signal on the Philips 'VLP' record are such that the optical scanning system must be able to perceive extremely fine details when the record is payeld. This requires an optical system with a large aperture and consequently a small depth of focus. It is also essential that the beam of light used for scanning should be centred very accurately on the information-carrying track on the record. The requirements specified here are about 1000 times more exacting than is usual in mechanical engineering. For example, the record surface may differ by 100-500 μm from a true plane while the depth of focus of the lens used to centre the scanning beam on the record is approximately 1 μm. The eccentricity of the track may amount to 100 μm, while the beam of light has to be kept centred on the track with an accuracy of 0.2 μm to avoid cross-talk [1]. To ensure continuous tracking of the information contained in a spiral of pits pressed in the surface of the record, the entire optical system has to be displaced on average by 2 μm radially for each revolution. Finally, a high-quality playback of the detected colour-television signal is only obtained if the speed at which the record rotates is constant to within at least 1 part in 10^3.

The 'VLP' record player is therefore provided with four control systems, which we shall describe below.

Focus control

Centrifugal forces that arise because the record is rotating at 25 rps can assist in flattening the record since its stiffness is relatively low. Mechanical vibrations are damped by the air layer between the record and the surface of the player and also by the internal losses in the disc material. Since a pressed record is never perfectly flat, axial displacements of the order of 500 μm are still encountered. The largest component of this displacement occurs at the fundamental frequency of 25 Hz corresponding to the speed of rotation. Measurement of higher-frequency harmonics showed that they decreased at a rate of 30-40 dB per decade. A control system used to keep the microscope objective focused on the record

Ir P. J. M. Janssen is with Philips Research Laboratories, Eindhoven, and Dr P. E. Day with the Philips Audio Division, Eindhoven.

surface to the required accuracy of 1 μm must therefore have a reduction ratio of at least 500 at 25 Hz.

The mechanical components of the control system have been kept as simple as possible. The microscope objective used for reading the video signal is suspended in springs and driven by a coil in a radial magnetic field in much the same way as a moving-coil loudspeaker (fig. 1). To restrict the energy dissipation in the region of the microscope objective a spring-mass system with an open-loop resonant frequency at 25 Hz was chosen.
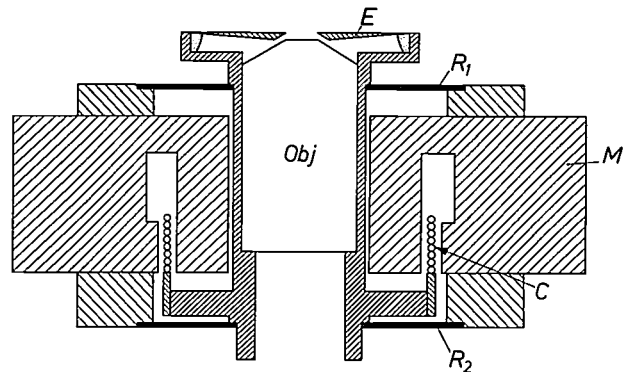
Fig. 1. The spring suspension of the microscope objective in the 'VLP' record player and the objective drive required for focusing the light beam on the 'VLP' record. Obj objective. R_1, R_2 springs. E measuring electrode. C drive coil. M magnet.

The distance between the objective and the record is determined by the capacitance between the metallized surface of the record and an electrode bonded to the objective (fig. 2). At a distance of 100 μm a capacitance measurement to an accuracy of 1% is sufficient to determine the position of the objective to within 1 μm.

This is not difficult to achieve, especially if the electrode is made so large that the effect of stray capacitance remains limited. The electrode must nevertheless not be too large because the nonlinear relation between the capacitance and distance would then give rise to excessive measuring errors, since the average distance could not be determined accurately if the electrode were inclined to the record surface. The size of the electrode also determines the resolving power of the measuring system, i.e. the highest spatial frequency that is still detectable. An electrode surface area of 1 cm^2, giving a capacitance of 10 pF, is found
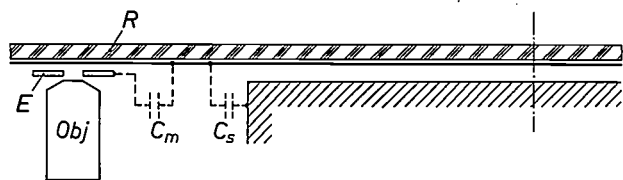
Fig. 2. The capacitive system for measuring the distance between the objective and the record. *Obj* objective. *E* measuring electrode. *R* the metallized 'VLP' record. *C* the capacitance of the electrode to the chassis of the 'VLP' record player. C_s is much larger than C_m and the total of the two capacitances in series is therefore almost entirely determined by C_m.

Radial tracking control

The beam of light used in the 'VLP' record player to scan the information has to follow the spiral track with an accuracy of at least 0.2 μm. Two control circuits are used for this purpose: one for continuous slow tracking of the spiral and the other for rapidly centring the spot on the track in spite of the eccentricity of the track. The first circuit will be discussed in the next section.

Because of the 100 μm eccentricity of the track on the rotating record the second control circuit has to
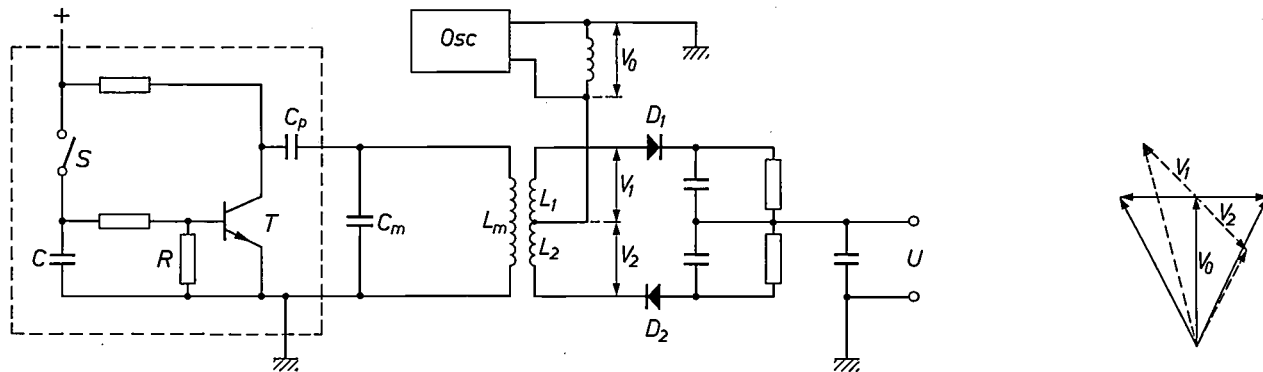
Fig. 3. The circuit for measuring the distance between the objective of the 'VLP' record player and the 'VLP' record. The oscillator *Osc* (output voltage V_0) is weakly coupled to the detector circuit formed by the measuring capacitor C_m and coil L_m. Voltages V_1 and V_2 induced in coils L_1 and L_2 will have a phase difference of 180°. Diodes D_1 and D_2 are given polarities such that the output signal at *U* is the amplitude difference between $V_0 + V_1$ and $V_0 + V_2$. When the resonant frequency of the circuit $C_m - L_m$ equals the oscillator frequency, V_1 and V_2 are both 90° out of phase with V_0, the two amplitudes are equal and no voltage occurs at the output. If the circuit is detuned, V_1 and V_2 will no longer be in quadrature with V_0 and a voltage will appear at output *U* whose polarity depends on the direction of detuning. This can be seen from the vector diagram, where the tuned condition is represented by continuous vectors and the detuned one by dashed vectors. The part of the circuit inside the dashed line is used to bring the objective slowly closer to the record. If switch *S* is closed, transistor *T* is conducting and capacitor C_p is connected in parallel with C_m. The resultant detuning of the circuit causes the objective to be moved away. When switch *S* is opened, *T* will gradually stop conducting, the rate of this change being determined by the value of *RC*. The objective will return at the same speed to the operating point determined by C_m and L_m.

to give a satisfactory compromise. The capacitance is measured with a circuit consisting of an oscillator and an FM ratio detector (*fig. 3*). Since the measuring electrode forms part of the detector circuit an unambiguous control signal is obtained over a large control range (1 cm - 10 μm). The oscillator frequency then determines the desired value of the capacitance and hence the distance between the electrode and the record.

Before putting a record on the player the objective is withdrawn to a safe distance from the record surface. To ensure that when the player is switched on the objective approaches the record smoothly and without overshooting, in spite of the relatively narrow bandwidth of the control loop a circuit is incorporated which enables the operating point of the control circuit to be shifted continuously from the retracted position to the operating point.

ensure a reduction of at least 500 at 25 Hz. The requirements demanded of the radial tracking system on account of the spectrum of the radial deviation are easily met in terms of the closed-loop control bandwidth. A larger bandwidth is required, however, since a rapid step response is necessary for perfect programmed scanning as employed for slow- and accelerated-motion pictures or finding a particular place on the record quickly.

To keep the spot centred on the track during playback, it can be moved radially by a slight movement of a pivoting mirror fitted immediately behind the objective in the light path. This mirror can pivot in a magnetic field and is rotated by a coil attached to it. The mirror-coil assembly is designed to be sufficiently rigid for the bandwidth required. The mirror and its

[1] See the article by G. Bouwhuis and P. Burgstede, this issue, p. 186.
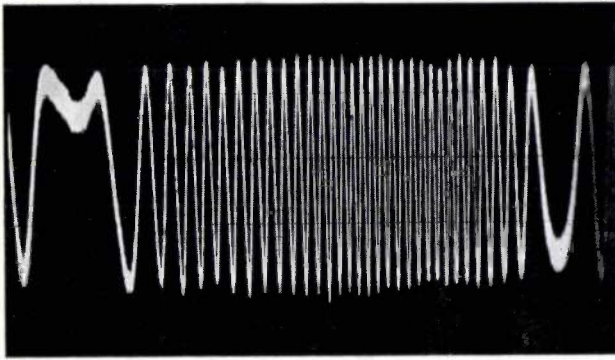
Fig. 4. Oscillogram of the tracking-control signal, recorded with the pivoting mirror stationary. Because of the eccentricity, a number of tracks now pass continuously over the two spots of light. The oscillogram shows the control signal for one half-cycle of this movement. It can be clearly seen that 32 turns pass the beam, so that the amount of eccentricity in this case was apparently 32 μm.

Fig. 5. Diagram showing different ways of using the picture recorded on the 'VLP' record. Twice per revolution, i.e. during the field flyback, an opportunity occurs to change from one turn to another. A reverse jump after every revolution yields a stationary picture (a), while a forward jump results in motion at twice true speed (b). A reverse jump after every second revolution reduces the speed by half (c), while a similar jump every half-revolution produces reverse motion at the correct speed (d).

suspension are designed as a unit to give a low frequency for the mechanical resonance. This, combined with the electrodynamic drive, ensures that a very low power is adequate for tracking.

The control signal is obtained with the aid of two auxiliary beams which are focused one on either side of the centre of the track. The light from each of these beams is reflected from the record surface on to a separate photodiode [1]. The difference between the two diode signals provides the signal that controls the pivoting mirror. The form which the signal takes is illustrated in fig. 4.

Fig. 5 shows how the television picture recorded on the 'VLP' record can be read out in such a way that accelerated, slow or even reverse motion can be obtained. This requires rapid movement of the beam from one turn of the track to another at the appropriate moment, i.e. during field flyback. To enable the beam to jump from one track to another during this period, use is made of the fact that the open control loop has a low resonant frequency, so that it behaves like a ballistic galvanometer. The beam jump is effected by opening the control loop and applying an accelerating current pulse, followed by an equally large retarding pulse through the mirror coil, and then closing the loop again (fig. 6).

The radial transport system

We have already seen that to enable the spiral on the 'VLP' record to be continuously tracked the entire optical system is displaced in the radial direction at an average rate of 2 μm for each revolution of the record. The whole optical system is therefore mounted on a carriage which moves along two guide rails [2]. Since the field of vision of the objective is 300 μm in diameter, an eccentricity of 100 μm will still leave

Fig. 6. Control of the pivoting mirror for pick-up of the picture. The figure shows from top to bottom, plotted as a function of the time t: the rectangular-pulse current applied to the coil and the resulting torque T, which is proportional to that current; the angular velocity ω of the mirror; and the position φ of the mirror. Between the two turns of the track the beam of light is insufficiently concentrated on either and no useful signal is obtained. In this region, however, the speed of displacement is high and the dwell time there consequently short.

100 μm, and consequently the positioning error may be about 50 μm. The radial velocity has to be 50 μm/s, which means that a first-order control system with a response time of 1 s will be satisfactory. The average current in the coil that operates the pivoting mirror is used as the control signal for the radial transport

[2] See fig. 2 of the article by K. Compaan and P. Kramer, this issue, p. 178.

system. The coupling thus formed between two control systems is obtained in the linear region of the control characteristic of the system for centring the beam on the track and there are therefore no complications arising from the interaction of the two systems.

Speed-control system

Special measures had to be taken to keep the speed of the 'VLP' record constant to within the required accuracy (at least 1 part in 10^3) since the mains supply cannot be expected to provide such a degree of stability and a simple mains-powered synchronous motor cannot therefore be used to drive the turntable.

Our solution may be briefly described as follows. The record is driven by a d.c. motor and a tacho-generator coupled to the turntable shaft delivers an a.c. voltage whose frequency is proportional to the speed of the shaft. By applying this signal to an RC network of very stable elements we obtain a voltage whose amplitude is a measure of the frequency and can therefore be used to control the speed of the d.c. motor. In other words, the RC network acts as a speed reference.

Summary. The 'VLP' record player is equipped with control mechanisms for focusing the optical system on the surface of the record, centring the scanning spot on the track, tracking, and keeping the speed of the record constant (25 rps). The focusing system has a reduction ratio of over 500 at 25 Hz and derives its error signal from a capacitive sensor. The objective drive is similar to that used in a moving-coil loudspeaker. The centring system also has a reduction ratio exceeding 500 at 25 Hz but here the error signal is obtained by optical means. The scanning spot is displaced by a small pivoting mirror which is driven by a rotating-coil system. For continuous tracking purposes a carriage moves the entire optical system radially, with the control system ensuring that the average deflection of the mirror remains small. The speed-control system (accuracy 25 Hz ± 0.1%) consists of a tachogenerator and a d.c. motor plus a precision-element network that converts the tachogenerator output into a control signal.

The Dutch national air-pollution monitoring centre at Bilthoven

The pictures on these pages illustrate the next important step towards the control of atmospheric pollution over the Netherlands. Regular readers of this journal will remember the description we gave some time ago of a monitoring network in the Rhine estuary region (Rijnmond) [1]. When this network, which was developed by Philips for the Rijnmond Authority, went into operation in 1969, it was still limited to monitoring the SO_2 concentration in the atmosphere, which is one of the most characteristic forms of industrial polution.

Since then work has continued in covering the whole of the Netherlands with a network of monitoring stations (eventually over 200), each of which will be connected by telephone line to the computer at a regional monitoring centre. The regional centres (there will shortly be nine of these) will be linked by telegraph lines to the national monitoring centre at Bilthoven. The map above shows the sites chosen for the 200 stations. The black dots represent detector stations, which are distributed over the country in a regular pattern, while the smaller open circles are extra stations added to the network because of local conditions. Regional centres are indicated by red dots and the national centre at Bilthoven by a red star. One completed part of the network went into operation in June 1973. For the time being only

the SO_2 concentration is being measured but in the near future it is intended to equip a number of monitoring stations with facilities for measuring the concentrations of other pollutants such as NO_2, NO, O_3, CO and H_2S. It will also be a fairly simple matter to add equipment for determining dust, hydrocarbons, fluorides, etc., in accordance with local needs. Forty of the monitoring stations are also equipped to measure meteorological parameters. The detectors are automatically calibrated every twelve hours on a signal from the regional monitoring centre. They require no attention apart from minor maintenance at three-monthly intervals.

A value measured by a detector is first converted into an electric current of between 0 and 20 mA. The current in turn is converted into a proportional frequency between 5 and 25 Hz. This is then used to modulate the amplitude of an audio-frequency carrier and the composite signal is finally conveyed by telephone line to the regional centre.

The job of the computer in each of the regional centres is to pre-process data from the monitoring stations, to pass them on in 'concentrated' form to a computer in the national monitoring centre and to control all stations in the region and the peripheral equipment in the regional centre itself. Every detector is scanned and the concentration values determined every minute and averages are computed every hour.

A dialogue teletypewriter in each regional centre records all reports of excess concentrations while a logging typewriter continuously registers air-pollution values and meteorological data. The same information can also be monitored continuously by an analog recorder. The dialogue teletypewriter can also be used at any time to obtain additional information from the computer and to feed into the computer the data required for calculations.

The regional monitoring centres in turn are controlled by the central computer at Bilthoven, which not only maintains communication with the regional computers but also processes, stores and presents information received from them and controls the peripheral equipment in the national centre. One item of peripheral equipment is a line printer which produces average values every hour. Threshold values can be programmed in the computer and if a signal exceeds the permitted value a lamp lights up on the map in the control room (see photograph) to indicate the monitoring station where the signal originated. Like the regional centres the national centre can ask at any time for additional information, which can be presented in a variety of ways (cathode-ray-tube display, logging typewriter, analog recorder, punched tape). The Bilthoven centre can therefore keep a close and continuous watch on the situation in all the regions and, if necessary, alert provincial authorities and industry, thus enabling the emission of pollutants to be restricted and an intolerable situation avoided.

The information stored can be used for statistical calculations, for the design of warning systems, etc., subjects in which the Royal Netherlands Meteorological Institute at De Bilt is also closely concerned.

[1] H. J. Brouwer, S. M. de Veer and H. Zeedijk, The SO_2 monitoring network in the Rhine estuary region, Philips tech. Rev. **32**, 33-41, 1971.

A simple model for alloys

A. R. Miedema

II. The influence of ionicity on the stability and other physical properties of alloys

Introduction

In the first part of this article [*] it was shown that a very satisfactory description of the energy effects encountered in the alloying of transition metals can be obtained from a model in which the basic starting point is the idea that the metal atoms in an alloy remain very similar to those in the pure metals. The heat of formation of the alloy arises because on one hand the chemical potential of electrons must become equal for both kinds of atomic cells, while on the other hand the discontinuity in the electron density (in electrons per cm³) that appears at the boundary of dissimilar atoms has to be smoothed out. This gives:

$$\Delta H = \mathrm{f}(c)\left[-\,Pe(\Delta\Phi^*)^2 + Q(\Delta n_{\mathrm{WS}})^2\right]. \quad (1)$$

Here ΔH is the formation enthalpy of the alloy, $\mathrm{f}(c)$ is a symmetrical function of the concentration of one of the metals and $\Delta\Phi^*$ and Δn_{WS} represent the discontinuity in electronegativity Φ^* and electron density n_{WS} at the boundary between dissimilar atomic cells; e is the electronic charge and P and Q are constants. Equation (1) describes the alloying behaviour of transition metals with one another and of transition metals with Cu, Ag, Au, Li, Ca and Sr. If transition metals are alloyed with trivalent or polyvalent non-transition metals, an extra negative contribution to the heat of formation appears, which is independent of which particular transition metal is alloyed with which particular p electron metal from the group Al, Ga, In, Tl, Sn, Pb, Sb and Bi:

$$\Delta H = \mathrm{f}(c)\left[-\,Pe(\Delta\Phi^*)^2 + Q(\Delta n_{\mathrm{WS}})^2 - R\right]. \quad (2)$$

The validity of this equation was demonstrated in Part I, and it was explained there how the parameters Φ^* and n_{WS} can be obtained. Values for the ratios Q/P and R/P of the constants in various cases can also be obtained from the analysis of the sign of the heat of formation.

In this second part we shall first be concerned with the absolute values of ΔH, that is to say we determine $\mathrm{f}(c)$ and the value of P, quantities that can be used to

calculate the heat of formation for any alloy of a transition metal. This is followed by a discussion of the heat of mixing of liquid metals, from which it can be concluded that equation (1) is also valid for (liquid) alloys of two non-transition metals.

Finally, the significance of charge transfer for the physical properties of alloys is examined. In the case of the heat of formation it was essential that the chemical potential of the electrons was made the same for the different atomic cells in an alloy. The transport properties, which are largely determined by the contribution of the d electrons to the conduction band, are strongly affected by this transfer of charge, particularly in the transition metals. In the present context these properties are of interest both in themselves, and also because they provide the information from which the quantitative relation between difference in electronegativity and charge transfer can be derived.

Absolute values of ΔH

In equations (1) and (2) another unknown apart from the constants P, Q and R is the dependence of ΔH on the concentration c. In addition, although values for Q/P and R/P were found in Part I, no value was found for P.

Let us start with $\mathrm{f}(c)$. For solid solutions the function would be expected to be $\mathrm{f}(c) = c(1 - c)$ if c is the concentration of one of the two metals, and size mismatch, Brillouin-zone effects and short-range order do not come into play. In ordered compounds no difference from statistical solid solutions would be expected provided that the concentration of one of the metals was small (again assuming that elastic energies associated with a difference in atomic size are relatively small; this is in fact almost always so if $|\Delta H|$ is fairly large — i.e. if $\Delta\Phi^*$ is large). This means that the heat of formation of ordered alloys in *fig. 14* should lie between curve *1*, which is the parabola $c(1 - c)$ for solid solutions, and the curves *2*, the tangents to the parabola at $c = 0$ and $c = 1$. An impression of the shape of $\mathrm{f}(c)$ can be obtained by comparing the heat of ordering with the total heat of formation for say 50/50 alloys possessing an

Dr A. R. Miedema is with Philips Research Laboratories, Eindhoven.

ordering temperature. These results suggest that on average the heat of ordering contributes about 30% of the total negative heat of formation.

It is also possible to estimate f(c) from experimental values of ΔH, provided that for a particular system there are a large number of ordered structures for which ΔH has been measured. These results have been plotted in fig. 14 for Ni-Al, Co-Al, Cd-Mg, Cu-Zn and the Pd-Al system, all systems whose measured ΔH values lie fairly symmetrically about the vertical $c = 0.5$. The results have been normalized with respect to the parabola $c(1 - c)$ by ensuring that at low concentrations the correct derivative is obtained. Close to the concentration 0.5 the difference between the solid curve and the parabola is about 30% of f(c); this agrees with the values for the relative magnitudes of the ordering energies.

In principle it is now possible to determine the parameter P by comparing experimental values for the heat of formation of alloys (and not just the sign) with equation (1) or (2), since the ratios Q/P and R/P have already been found. Little experimental information is available about the absolute value of ΔH for cases in which $\Delta\Phi^*$ or Δn_{WS}, and hence ΔH, are large. *Fig. 15* shows all the ΔH values that clearly differ from zero and are related to the groups discussed earlier (two transition metals, transition metals with p metals and hydrides). To find P, ΔH is plotted against the quantity f(c)$[-e(\Delta\Phi^*)^2 + (Q/P)(\Delta n_{WS})^2 - R/P]$, making use of the values found earlier for Q/P and R/P; R is of course put equal to zero for the hydrides and alloys of two transition metals. The amount of data available is too small to be able to determine whether P has a different value for the three groups of alloys. When all the results are taken together the mean straight line corresponds to $P = 0.85$ V^{-1}. A detailed table has been shown elsewhere [9] in which the ΔH values calculated with this value of P for all the alloys quoted have been compared with experiment.

Alloys of two non-transition metals. Liquid metals

To what extent do the model, and with it equation (1), also describe the alloying behaviour of non-transition metals? It is known that for example in the compounds or elements with the diamond-type structure (GaSb, InAs, Si, Ge) a large contribution to the energy originates from Brillouin-zone effects. The presence of eight valence electrons per unit cell makes these 'metals' semiconductors. The relatively large energy gap at the Fermi surface causes a fairly large decrease in energy. An impression of the magnitude of this energy contribution can be obtained from the latent heat of fusion, which is considerably larger for the four semiconduc-
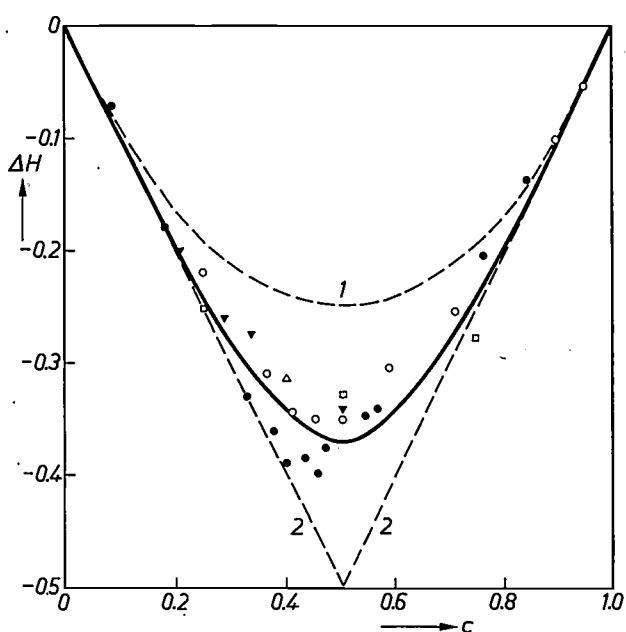
Fig. 14. The heat of formation ΔH of an ordered alloy (or compound) of two metals as a function of the atomic concentration c. Curve 1 represents the parabola that would be expected for solutions in which the metal atoms are statistically distributed over the lattice sites. The curves 2 give a lower limit for the factor f(c) of equation (1) in ordered alloys. O Ni-Al, □ Cd-Mg. ▼ Co-Al, △ Cu-Zn, ● Pd-Al. The data have been taken from the publications of notes [16] and [17].
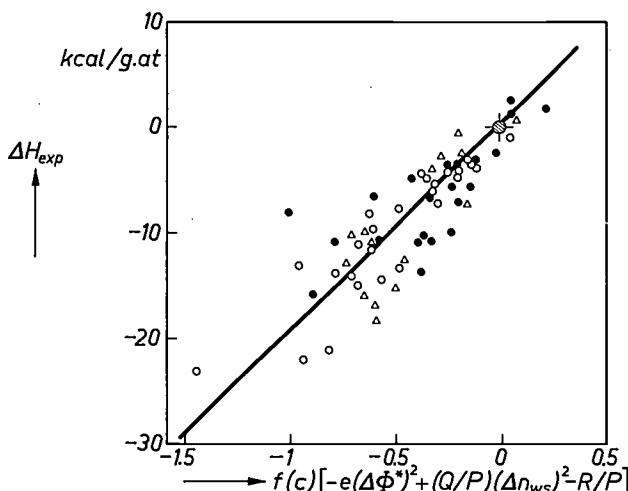
Fig. 15. Determination of the parameter P in the ionic energy contribution $-Pe(\Delta\Phi^*)^2$. The value of P can be derived from the slope of the straight line through the origin giving the best fit to the plotted points. ● Alloys of two transition metals. O Alloys of transition metals with non-transition metals of valency greater than 2. △ Hydrides of transition metals.

[*] A. R. Miedema, A simple model for alloys, I. Rules for the alloying behaviour of transition metals, Philips tech. Rev. 33, 149-160, 1973 (No. 6).
[16] R. Hultgren, R. L. Orr, P. D. Anderson and K. K. Kelley, Selected values of thermodynamic properties of metals and alloys, Wiley, New York 1963.
[17] M. Ettenberg, K. L. Komarek and E. Miller, Metallurg. Trans. 2, 1173, 1971.

tors mentioned above than for ordinary metals, whose latent heats of fusion are usually all very similar. The contribution of Brillouin effects to the heat of formation can be estimated from the unusually large latent heat of fusion. It is clear that this contribution only becomes comparable in magnitude with the terms of equation (1) for crystals with the ZnS or NaCl structure.

In liquid alloys there should in principle be no Brillouin-zone effects. The equation

$$\Delta H = c(1 - c) \left[-Pe(\Delta \Phi^*)^2 + Q(\Delta n_{WS})^2 \right] \quad (1')$$

would therefore be expected to apply for two non-transition metals.

In the first article of note [9], which presented an analysis of liquid metals with $\Phi^* = \Phi$, this relation has already been fairly clearly established. It is interesting to note that (1') is very similar to the equation used by N. F. Mott to predict (to an accuracy of 80%) whether two liquid metals were miscible. Mott used the electronegativity X, which we know to be approximately proportional to Φ^*, and a positive term proportional to $[\Delta(E_v/V_m)^{\frac{1}{2}}]^2$, where E_v is the latent heat of evaporation of a metal. The parameter n_{WS} used here represents $[\Delta(KV_m)^{-\frac{1}{3}}]$, and since the compressibility K is approximately inversely proportional to the latent heat of evaporation this means that Mott's equation is almost identical with equation (1'). Comparing the absolute values found with the two equations shows that for liquid metals Q is relatively a little higher than we found for the solid transition metals. The conclusion is that the difference in electronegativity and in electron density are also responsible for the greater part of the heat of mixing in non-transition metals. The size mismatch between the atoms of an alloy also contributes to some extent to the positive terms in ΔH (which can lead to decomposition in the liquid phase), but this contribution is considerably smaller than has recently been suggested by B. Predel and H. Sandig [18]. It is perhaps fairly obvious that energy effects depending on a difference in atomic size could be confused, for disordered solid and liquid alloys, with effects due to difference in density. Metals that differ considerably in electron density at the boundary of the atom will in general also differ considerably in atomic radius.

Magnetic properties and charge transfer

In the model presented here it is of essential importance that the negative term in the heat of formation of alloys is concerned with the transfer of electrons, and hence with ionicity. How large is the charge transfer in a particular case, expressed for example in numbers of electronic charges per atom? It is also interesting to consider to what extent charge transfer affects other properties of alloys besides the heat of formation.

In our model, in which the negative part of ΔH is determined by f(c) and $\Delta \Phi^*$, the charge transfer per atom will also be a function of only c and $\Delta \Phi^*$. If Φ^* does indeed represent the value of the chemical potential for electrons in the correct units, then the ionic contribution contained in ΔH is equal on one hand to $-P$ f(c) $(\Delta \Phi^*)^2$, but we also have:

$$\Delta H = -c_A e \, \Delta Z_A \, \Delta \Phi^*/2. \quad (3)$$

The quantities c_A and ΔZ_A represent the concentration of metal A and the charge per atom A in the alloy. The equation states in fact that the total charge displaced, $c_A e \Delta Z_A$, is on average transferred over the half potential difference, $\Delta \Phi^*/2$. (During the transfer of charge this potential difference is gradually reduced from $\Delta \Phi^*$ to zero; the average potential difference was $\Delta \Phi^*/2$.) From the above it follows that

$$\Delta Z_A = 2 \, \Delta \Phi^* \, f(c) \, P'/c_A. \quad (4)$$

We now distinguish between P' and P, since it is not certain that the Φ^* scale does in fact give the chemical potential in the correct units. We do know that our scale is a good relative measure of the chemical potential, but in the procedure selected it is by definition normalized to the scale of the work function Φ. If however there is a proportionality constant relating Φ and Φ^*, then it will also apply to the quantities P and P'.

For solid solutions f(c) is equal to $c_A(1 - c_A)$, and (4) therefore becomes:

$$\Delta Z_A = 2 \, \Delta \Phi^* (1 - c_A) P'. \quad (5)$$

The value of P' can be derived from experiments in which ΔZ_A manifests itself in some other way than a reduction of energy. An example for the transition metals is given by the *magnetic* properties.

In fig. 7 (Part I) the three groups of the 3d, 4d and 5d transition metals were quoted together, with the appropriate values of Φ^*. The partially filled d shell indicates that for these metals the d levels make a considerable contribution to the density of states at the Fermi surface. This is associated with a relatively large contribution to the linear term in the specific heat of metals and a large contribution to the temperature-independent (Pauli paramagnetic) susceptibility (for 4d, 5d and some 3d metals) or the ferromagnetic behaviour (for other 3d metals).

The charge-transfer effect can have a considerable effect on the magnetic behaviour of alloys. In the alloy Pd_3Y — formed from the last and the first transition metal of the 4d series — the difference in Φ^* is very large. The Y gives up electrons, the Pd receives electrons. The effect is as if Y displaces itself to the left in

the periodic system, thus losing its transition-metal character. The Pd displaces itself to the right, the d levels become completely filled and Pd becomes diamagnetic, like Ag. As a result the alloy Pd_3Zr does not have the properties of a transition metal (the high susceptibility). This change in properties when the number of d electrons becomes very small or is increased to ten allows an estimate of charge transfer to be made. The magnetic susceptibility of solutions of Zr in Pd decreases sharply with increasing Zr concentration and appears to go to a 'diamagnetic' value for 12.5 at. % Zr (the compounds Pd_4Zr and Pd_3Zr are also diamagnetic). From equation (5), and assuming that pure Pd metal has a number of 0.35 holes in its d level then it follows from the diamagnetic nature of $Pd_{87.5}Zr_{12.5}$ that $P' = 0.55$ V^{-1}. Comparable values for P' also follow from the magnetic properties of alloys of Co and Ni with the transition metals of the first two columns, if again it is assumed that the absence of ferromagnetism in intermetallic compounds of Co and Ni at $T = 0$ indicates an almost completely filled d level. On average we find that $P' \simeq 0.6$ V^{-1} (see note [9]).

Charge transfer can also be estimated from experiments on X-ray fluorescence (A. Wenger et al. [19]). Wenger obtained a value for the change in the number of d electrons at the Fe and Co sites in compounds with Al. The value of P' in this case is again in the region of 0.6 V^{-1}.

Note that the value found for P' corresponds to a relatively large charge transfer. In a disordered AB compound with $\Delta \Phi^* = 1$ V equation (5) shows that the charge is given by $A^{-0.6}B^{+0.6}$; here it should not be forgotten that on one hand still greater values of $\Delta \Phi^*$ can be found in practice, while on the other hand ΔZ is larger by a factor of 1.5 in ordered AB alloys (i.e. compounds).

The value found in this subsection for P' (0.6 V^{-1}) differs from the value found for P (0.85 V^{-1}) when the calculation is based on energy data alone. As was noted earlier, this does not necessarily mean that the model is 'wrong', but it can be taken as an indication that the scale of Φ^* ought really to be stretched by a factor of 1.4 to enable the internal chemical potential to be expressed in the correct units. On further consideration this result suggests that a proportionality exists between the work function Φ of a metal, the discontinuity D in potential that is present because there is an electric-dipole layer at the surface of a metal, and the chemical potential Φ^*, which relates to the interior of a metal. The difference between P and P' suggests that

$$\Phi : D : \Phi^* = 1 : 0.4 : 1.4$$

could be a general relation for all metallic elements.

Ionicity and the physical properties of alloys

In the previous section the initial assumption in seeking quantitative data for the relation between the charge transfer per atom and the difference in electro-negativity was that it was possible to derive information about the ionicity not only from the energy effects but also from the magnetic properties of alloys, for example. In fact, a basic assumption is involved here. In the model, as it was used in considering heats of formation, it was assumed that the distribution of charge (electrons) in the crystal lattice of an alloy could be found by first taking the charge distribution for the two kinds of pure metal and then applying an important correction (making Φ^* equal) to this distribution. In the description of magnetic properties we have gone much further. Here it has been assumed, for example in the case of the Pd-Zr alloys, that in the resulting distribution obtained for the alloy the original metals would still be clearly distinguishable. To give a picture of the situation it is assumed that the electron states (wave functions) of the alloy can be represented to a good approximation as a combination of Pd-like states and Zr-like states. In the alloy, in this explanation of the diamagnetic nature of $Pd_{87.5}Zr_{12.5}$, Pd has obtained a completely filled d level while that of Zr has become completely empty, so that the resulting metal has lost its transition-metal nature.

It is by no means obvious that electron states in an alloy can be obtained by adding metal A-like states and metal B-like states together, although this picture has been fairly widely used in descriptions of magnetic properties of Pd, Ni or Co alloys. No theoretical justification for this has been given; a possible qualitative explanation could be that the 'overlapping' of electron states of neighbouring atoms is relatively small in a lattice of transition-metal atoms, i.e. n_{WS} is small compared with the average density of d electrons in het atomic cell. A metal can be thought of as being made up from a number of free atoms that are first located a long way apart and then gradually approach one another. In this picture transition metals are metals in which the original free atoms have not yet come very close together. The 'overlap' still remains small; on this basis it might be expected that the properties of free atoms (and in alloys those of the *differing* atoms) could be recognized in the metal. In non-transition metals, on the other hand, n_{WS} is not small compared with the average d electron density and the atomic nature is lost.

In spite of the doubt, for theoretical reasons, about an atomic description of the physical properties of

[18] B. Predel and H. Sandig, Z. Metallk. **60**, 208, 1969.
[19] A. Wenger, G. Bürri and S. Steinemann, Solid State Comm. **9**, 1125, 1971.

alloys of transition metals, this kind of description has been found to give surprisingly good results in practice. For example, it has been found possible to give a simple formula for the specific heat at low temperature γ — which is a measure of the density of electron states at the Fermi surface — for solid solutions of transition metals. It is assumed in the first place that γ is atomic in nature for transition metals: γ is mainly determined by Z, the number of valence electrons per atom. For an alloy of two metals A and B it is assumed in the second place that

$$\gamma_{\text{all}} = c_A \gamma_A + c_B \gamma_B. \tag{7}$$

Here γ_A is the γ value for the A atoms in the alloy, and γ_B is the γ value for the B atoms; γ_A and γ_B depend on the number of valence electrons possessed by A and B in the alloy. This number is found for each type from the number Z_0 that applies for the pure metal, and the electron transfer ΔZ in the alloy:

$$Z = Z_0 + \Delta Z. \tag{8}$$

For Ti, Zr and Hf Z_0 is equal to 4, for V, Nb and Ta Z_0 is equal to 5, for Cr, Mo and W Z_0 is equal to 6, etc. (fig. 7, Part I). The value of ΔZ follows from equation (5) with $P' = 0.6 \text{ V}^{-1}$. Before we can use the formula, we still need to know the relation between γ and Z. This is given in *fig. 16*. The curve goes through the points for the pure elements, and is drawn in such a way that all the known data for paramagnetic transition metals are well reproduced. *Fig. 17* shows that with the simple formula a good agreement is obtained between calculated and measured values for a large number of alloys.

A similar atomic description also serves for the transition temperature for superconductivity in solid solutions of two transition metals. The assumption that electron states of metal A and of metal B can be added together corresponds to the addition of $[\ln(T_c/\Theta_D)]^{-1}$ for those metals in the theory of superconductivity. The transition temperature T_c is measured with respect to the Debye temperature (Θ_D) for the lattice vibrations. Equation (7) now becomes:

$$\left[\ln(T_c/\Theta_D)\right]^{-1}_{\text{all}} = c_A \left[\ln(T_c/\Theta_D)\right]^{-1}_A + c_B \left[\ln(T_c/\Theta_D)\right]^{-1}_B. \tag{9}$$

The characteristic function that represents the dependence of T_c/Θ_D on Z is shown in *fig. 18*, in which T_c/Θ_D is plotted to give a linear curve for $\ln(T_c/\Theta_D)^{-1}$. Although equation (9) is slightly more difficult to use than equation (7), the 'recipe' for finding the T_c of alloys of transition metals is really quite simple. The results obtained on comparing the calculated and measured values are again surprisingly good [9]. A
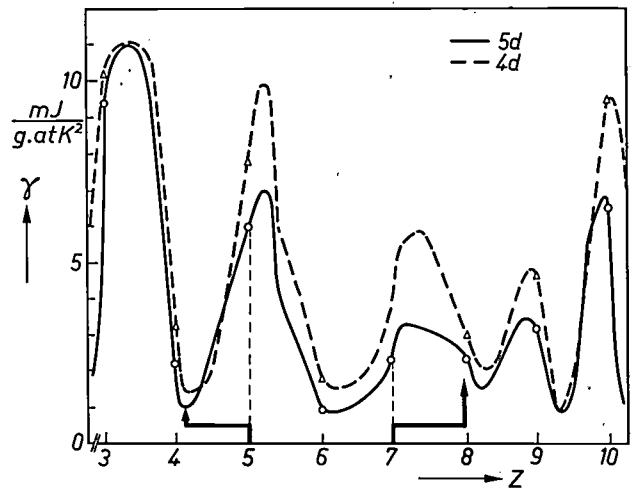
Fig. 16. The coefficient γ of the linear term in the specific heat of metals, plotted against the number of valence electrons per atom Z for the 4d and 5d metals. To give an example of the way in which equation (7) is applied the figure shows the Z values for Ta and Re in an alloy Ta$_{50}$Re$_{50}$ that does not in fact exist as a solid solution. The value Z_0 of Z for the metals in the pure state is 5 for Ta and 7 for Re.

simple demonstration of the applicability of the atomic curve for T_c/Θ_D as a function of Z shown in fig. 18 can be found in the change in T_c of Nb in alloys with small quantities of a second metal. As fig. 18 shows, pure Nb, for which $Z_0 = 5$, lies at a point of the curve at which the derivative of T_c/Θ_D with respect to Z has a high value. If a metal is dissolved in Nb, then it is relatively unimportant that the second metal gives a characteristic contribution to $[\ln(T_c/\Theta_D)]^{-1}$ different from that
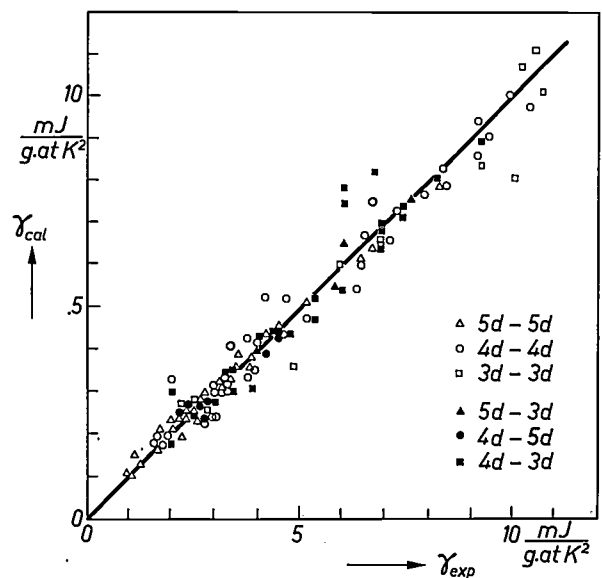
Fig. 17. Comparison of experimental data on γ for solid solutions of two transition metals with the values γ_{cal}, calculated from equation (7). The points lie close to a straight line of slope 1. The figure shows which types of alloy the various kinds of points in the figure refer to.
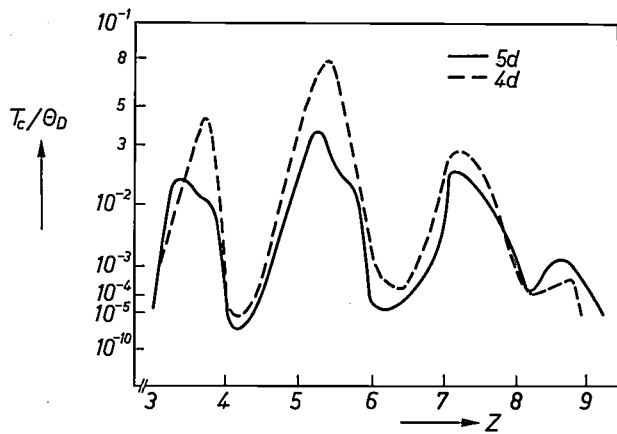
Fig. 18. The ratio T_c/Θ_D of the temperature at which a transition metal becomes superconducting and its Debye temperature, plotted against Z, the number of valence electrons per atom. The solid curves apply for 5d metals, and the other curves for 4d metals.

given by pure Nb. However, it is important that the dissolved metal can change the Z of the Nb. If the dissolved metal takes up electrons from the Nb (if it is more strongly electronegative than Nb, higher Φ^*) then T_c/Θ_D will decrease; if the dissolved metal gives up electrons, then T_c will be able to increase because of the high value of the derivative of T_c/Θ_D with respect to Z at the location of Nb on the curve. In this argument it would be expected that the change in T_c for niobium on the solution of small concentrations of a second metal would chiefly be determined by $\Delta\Phi^*$ between Nb and this metal. The truth of this prediction is shown in
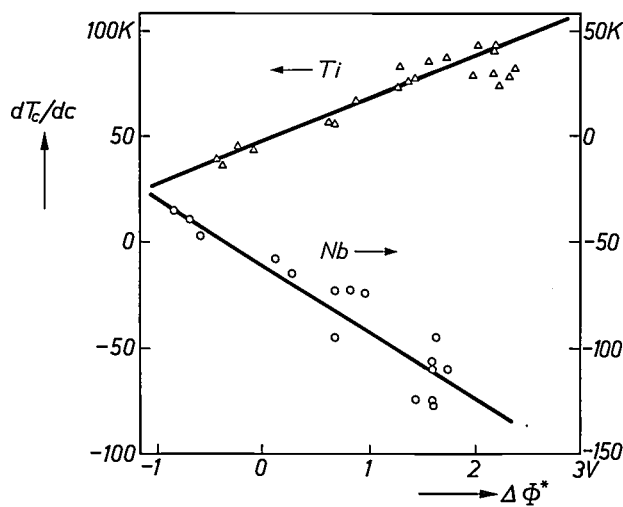
Fig. 19. The derivative of the transition temperature for superconduction T_c with respect to the concentration c, plotted against the difference in electronegativity $\Delta\Phi^*$ for diluted Nb [11] and Ti alloys [20]. (The Ti alloys had 5% of Rh added to stabilize the b.c.c. structure.) For each of these alloys the points lie very nearly on a straight line; the line for the Ti alloys rises for increasing $\Delta\Phi^*$, the other one falls.

fig. 19, in which dT_c/dc, the change in T_c with the concentration of metal added, is plotted against $\Delta\Phi^*$. The metals with the higher values of Φ^* reduce T_c very noticeably; the relation between dT_c/dc and $\Delta\Phi^*$ can be approximated quite well by a straight line. Fig. 19 also shows the change in T_c for Ti in alloys, plotted against $\Delta\Phi^*$ for the dissolved element. The slope of the curve of T_c/Θ_D as a function of Z is also relatively large for Ti ($Z_0 = 4$), but the sign is different.

As was noted at the beginning of this section, there is no obvious reason for using an atomic description for alloys of non-transition metals; we can only say that it does appear to describe energy effects. It is consequently more difficult to justify the use of Wigner-Seitz cells for these alloys. A definition that appears reasonable for the boundary between neighbouring atomic cells in a transition metal is the minimum in the electron density as a function of location. This is no longer possible in non-transition metals, and it will be obvious that it is difficult to speak of the transfer of electrons while it is not in fact possible to define the location of the boundary between atoms A and atoms B. It is also much less obvious which physical properties will be affected by the transfer of charge. Whether $Na^{+0.5}K^{-0.5}$ or $Na^{-0.5}K^{+0.5}$ is written as the formula for an Na-K alloy does not affect the metallic properties at all. Similarly, it makes no obvious difference to the semiconductor properties of GaSb whether we are dealing with Ga^+Sb^-, GaSb or Ga^-Sb^+. However, the existence of charge transfer can be observed indirectly in the change in total volume on the alloying of two non-transition metals, from the appearance of electric-field gradients (which can for example be studied at the locations of atomic nuclei in nuclear-resonance experiments), or from measurements of the isomer shift in Mössbauer-effect experiments. Let us consider Ag and Au in such an alloy of two non-transition metals. The two metals both have the f.c.c. structure and very nearly the same atomic volume. Nevertheless, a fairly large isomer shift is found in Mössbauer experiments on Ag-Au alloys. This shift is a direct measure of the probability of s electrons arriving at the location of the nucleus of Au. While it is true that it is not (yet) possible to convert a change in isomer shift into a transfer of charge, it is indeed possible to see that the ionicity does have consequences for the physical properties, and not only for the transition metals.

To summarize quickly the results of the second part of this article, we see that in alloys of two transition metals the negative contribution to the heat of formation is almost entirely determined by the difference in electronegativity $\Delta\Phi^*$. Difference in electronegativity

[20] Ch. J. Raub, Z. Physik **178**, 216, 1964.

signifies transfer of charge. This is large: ΔZ is proportional to $\Delta \Phi^*$, and for solid solutions the proportionality constant is given by $1.2 (1-c)$. Physical properties of transition metals can be particularly sensitive to electron transfer, as has been shown in this article for the contribution of the electrons to the specific heat and the transition temperature T_c for superconduction.

Summary. This part of the article first treats the influence of the concentration on the heat of formation of alloys and the value of the coefficient of the electronegativity contribution, from which the absolute value of the heat of formation can be calculated. It is found that the relation derived in I for the heat of formation of alloys of two transition metals is also valid for liquid alloys of two non-transition metals. Difference in electronegativity signifies transfer of charge, to which the physical properties of transition metals can be very sensitive; this is demonstrated for the electron contribution to the specific heat and for the temperature at which the alloy becomes superconducting.

The development of field-effect transistor electronics

F. M. Klaassen

The text of the article below is an abridged and slightly modified version of the address given by the author on 8th December 1972 on his inauguration as Professor Extraordinary in Semiconductor Electronics at Eindhoven University of Technology. In his address, in which he outlined the history of the bipolar transistor and the older field-effect transistor, Prof. Klaassen gave particular attention to the recent increase in importance of the field-effect transistor, after years of obscurity, in digital integrated circuit applications.

Almost exactly twenty-five years ago, in December 1947, John Bardeen and Walter H. Brattain were attempting to obtain electrical amplification by applying the well known field effect to the reverse current through a germanium point-contact diode when they observed a change in the current that was not entirely what they had expected [1]. This result was later found to have been obtained because the gold field-effect electrode was not, as had been intended, separated from the germanium by a thin film of oxide, but was in fact in direct contact with the semiconductor. They came to the conclusion — which later turned out to be the right one and for which they received the Nobel Prize — that the gate electrode injected into the germanium an additional mobile charge that was accumulated by the point contact. This marked the birth of the first practical transistor. Although it was the unexpected result of a physical experiment, this transistor was also the outcome of a programme of investigations that had already been in progress for about ten years, with the aim of finding a solid-state analogue for the thermionic valve. Such an analogue was particularly desired by the telephone industry, which, with the expansion of telephone networks, wanted a device less greedy of power and longer-lived than the thermionic valve.

If we look at a photograph of the primitive construction of the first transistor (*fig. 1*), which gives a striking image of a nearly bygone experimental era, it is clear why very few people saw this invention as a device that would revolutionize electronics and help to change the face of society, and that it would do this in record time. Spurred on by the best and the worst characteristics of free enterprise and by the personal ambition of the experimenters, and resting on the understanding of
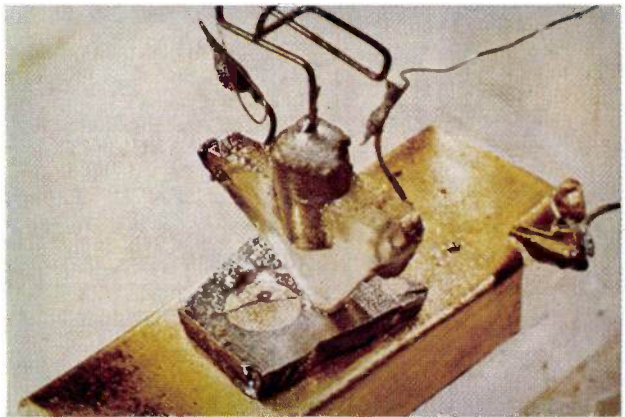
Fig. 1. The transistor produced by J. Bardeen and W. H. Brattain [1].

the solid state gained in the steady struggle of research into unrelenting matter, semiconductor technology has undergone in twenty-five years an evolution that is scarcely rivalled in the history of technology, and with what a result! Without transistor technology there would be no computer industry, no moon-walks, no communication or weather satellites and presumably no thousands of millions of portable radios, no hundreds of millions of telephone subscribers, to mention just a few spectacular applications. Bardeen and Brattain's discovery proved to be a key to the 'all at once world'.

In electronic engineering itself the change brought about was equally dramatic. The advent of the integrated circuit brought an end to the old three-way split that had long dominated electronics. There is no longer a place for the independent component designer, the independent circuit designer, or the independent sys-

Prof. Dr F. M. Klaassen is with Philips Research Laboratories, Eindhoven and is Professor Extraordinary in Semiconductor Electronics at Eindhoven University of Technology.

[1] J. Bardeen and W. H. Brattain, The transistor, a semiconductor triode, Phys. Rev. **74**, 230-231, 1948.

tems engineer. Since there are very few who can be masters in all three disciplines, advances in these fields will only be made by interdisciplinary cooperation. Such cooperation requires at the very least, however, an understanding of the other man's language or, better still, an understanding of the other man's often intuitive basic knowledge.

It was a logical consequence of the discovery in 1947 that the dominant role in the semiconductor realm in the period that followed has been played by the bipolar transistor, a device whose operation depends fundamentally on the injection and diffusion of electrons and holes, and whose behaviour therefore deviates completely from Ohm's law. The field-effect transistor received much less prominence, even though this transistor was invented nearly 20 years before 1947, is simpler in structure and is more closely related to a pure resistance in its behaviour. It is only in recent years that it has begun to look as if the bipolar transistor will have to give up much of the ground it has gained. This late emergence of the field-effect transistor, after being so long overshadowed by the bipolar device, seems interesting enough for closer examination.

Since the reader will probably be fairly familiar with the history of the field-effect transistor [2], it should be sufficient here to recall simply that one of the last major steps was the replacement of aluminium as gate material by polycrystalline silicon (*fig. 2*), to obtain a sufficiently low threshold voltage [3].

It was this last technique, not spectacular on the face of it, but in fact one that required a radical change of process, that initiated the rapid advance of the MOST a few years ago. One of the contributory factors here was that a reliable double-layer wiring pattern fits fairly naturally into this process; another was that the transistors can be packed closer together. A technology that is also of interest besides the now current MOS diffusion technology is ion implantation [4], in which very thin layers of the required conduction type are implanted in the silicon surface by bombarding this surface with fast ions, making it possible to preset the threshold voltage. Transistors with positive and negative threshold voltage can then be made simultaneously in one crystal chip, an arrangement for electronic circuits that is very attractive. Indeed it has already been found possible to make the threshold voltage so low that millivolt logic is a feasible proposition. This brings us a long way in the direction of the extremely low energy consumption that the human brain needs to perform logical processes.

As to the usefulness of the field-effect transistor in electronics, in the first place it is of course true that there was no dramatic increase in its use during the
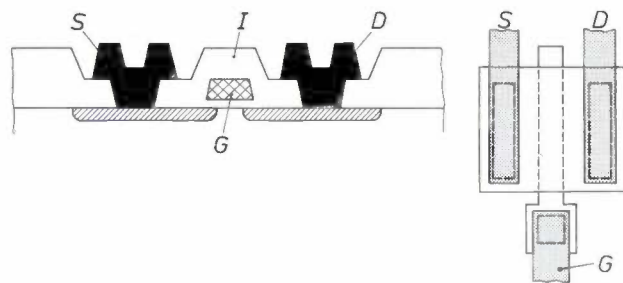
Fig. 2. Schematic cross-section of a field-effect transistor with a gate of polycrystalline silicon. *S* metal connection of the source. *D* metal connection of the drain. *I* insulating oxide layer. *G* gate.

first ten years. In itself this is not so surprising, for there is often a considerable time lag between the conception of an idea and its application. The applicability of an idea can generally be demonstrated, but a broad application is often a problem because of the lack of adequate background information, such as a complete understanding of the physical operation, knowledge of the properties of materials used and, last but not least, process control. Progress in fits and starts is therefore rather the rule than the exception. Nevertheless the slow growth of field-effect transistor electronics can be partly attributed to trivial, or perhaps, realistic causes. The price limited applications to the strictly professional sphere for a long time; the MOSFET (MOST for short) had a bad reputation for its disastrous oversensitivity to electrostatic charges; and most important, the bipolar transistor, which had led the field for ten years, had set a particular style for electronic design, so that the FET was hardly ever used as a pure substitute for the thermionic valve. The junction FET did better than the MOSFET, however, and in ten years it gradually grew to its present total market of eighty million devices a year. Because of its favourable combination of reasonably low-noise amplification with a fairly large bandwidth, the junction FET proved to be exceptionally suitable as a pre-amplifier for signals originating from sources of high internal impedance, and was thus able to fill a gap that the bipolar transistor could not. This low noise level is very closely related to the absence of input-current noise in the junction FET, which is not the case in the bipolar transistor. Of course, not everyone is satisfied. For example, designers of preamplifiers for television cameras, radiation detectors and operational amplifiers are constantly demanding a lower noise level. The ability to meet that demand implies on one hand greater process control to reduce the low-frequency noise — whose nature is not yet entirely understood — and on the other hand it implies refined masking techniques to fabricate shorter FETs for a lower i.f. noise. The high hopes of being able to use

junction FETs for the first stage in telecommunication receivers remained largely unrealized. The very low input cross-modulation, compared with that of the ordinary transistor, which was expected on account of its square-law characteristic, has been found unattainable in practice, for one thing because of the need to transform the signal from the (low impedance) aerial to the correct level for the impedance of the junction FET input. There are also indications that electron-velocity saturation in high-transconductance transistors has an unfavourable effect on the distortion.

Because of requirements that are very nearly irreconcilable, it is very difficult to incorporate the junction FET in an integrated circuit. For instance, a thin high-conductivity channel is required for a junction FET with a high transconductance, but this requirement is positively disastrous for the breakdown voltage of a bipolar transistor included in the circuit. Hardly anyone has yet succeeded in designing good integrated operational amplifiers with a junction FET input. To match the two FETs of the differential input stage the transistors usually have to be twisted around the chip like a pair of snakes!

A climax in the development of the junction FET was reached some time ago when a field-effect device surpassed the power amplification performance of the bipolar transistor in the microwave region (fig. 3). An interesting aspect in the achievement of this record is that silicon, the dominant material in transistor manufacture, was replaced here by gallium arsenide, because this material has a higher electron mobility and thus a shorter electron-transit time [5]. Although gallium arsenide has certain properties that make it less suitable for the fabrication of ordinary transistors, these do not matter so much FETs. Whether this development will lead to large-scale application is doubtful, however, since in practice the bandwidth of a FET amplifier is much smaller than would be expected from the cut-off frequency given in fig. 3, because of its high impedance.

The use of the MOS transistor in linear electronics has largely been a failure; but with one exception: the ideal insulation between gate and channel makes the MOST an unrivalled electrometer. MOST integrated circuits have not, however, been able to replace valve circuits directly. This is very easily explained. Although the combination of one MOST as an active element in series with another MOST as the load forms an ideal amplifier which in principle takes up less chip area than its bipolar counterpart, a difficulty arises at the input and output of the circuit, where only transistors of relatively large dimensions can operate. At the input, this need for large size is connected with the signal-to-noise ratio, the drift and the ability to with-
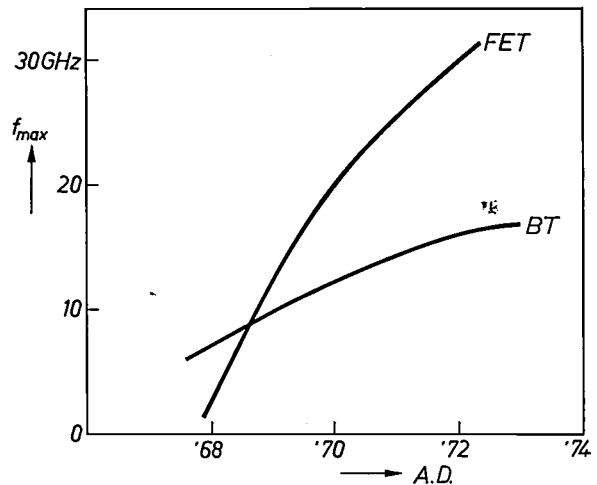
Fig. 3. Increase in maximum useful frequency of bipolar and field-effect transistors in recent years. The field-effect transistor (curve FET) took over from the bipolar transistor (curve BT) some years ago.

stand interfering electrostatic charges, whereas at the output the dimensions have to be large to obtain a high transconductance at higher current. In the output-transistor the very much poorer ratio between transconductance and current — 1 or 2 compared with 40 for a bipolar transistor — also creates difficulties. An all-MOST circuit will only be attractive for circuits with a large number of internal stages, but linear circuits are not usually so large.

Finally, a word about the noise level of the MOST, which is relatively rather high, particularly at low frequencies, and is usually about ten times greater than that of the FET. Although this effect has not yet fully been explained, there are nevertheless some indications that the noise is due to a slow interaction between the free charges in the channel and traps situated in the silicon-oxide interface [6]. Although progress in passivation has brought with it a gradual reduction of the noise level, the decay time involved has certainly thinned out the supporters in the course of time.

The picture, however, is not an entirely gloomy one. Indeed, it has recently come to appear that MOST integrated circuits have a great future ahead of them in the field of large logic circuits.

What is the reason for this breakthrough, and why has it come so late? The answer is evident; the use of MOST technology leads in many cases to lower system

[2] See the introduction to the article by H. C. de Graaff and H. Koelmans in Philips tech. Rev. 27, 200, 1966.
[3] See for example J. A. Appels, H. Kalter and E. Kooi, Philips tech. Rev. 31, 225, 1970.
[4] See for example F. Faggin and T. Klein, Silicon gate technology, Solid-State Electronics 13, 1125-1144, 1970.
[5] W. Baechtold, W. Walter and P. Wolf, X and Ku band GaAs M.E.S.F.E.T., Electronics Letters 8, 35-37, 1972.
[6] F. M. Klaassen, Characterization of low 1/f noise in MOS transistors, IEEE Trans. ED-18, 887-891, 1971.

costs, and in the second place the permissible dissipation imposes less of a restriction on the number of transistors to be integrated. These facts are sufficiently important for us to go somewhat deeper into a number of aspects.

· One of the basic circuits with which all logic operations can be carried out is the MOST inverter [7]. This consists of two very small transistors in series, one of which functions as an active switching element and the other as a passive load (*fig. 4*). Compared with bipolar circuits, which are often no more than IC translations of discrete-component solutions, a combination of this type is not simply a more advanced form of integration; it is also more compact, since not a single resistor is required anywhere in the circuit. In bipolar ICs it is the resistors that are the expensive items since, particularly in circuits with low dissipation per gate — a first prerequisite in large-scale integration — the resistors have to have a high value, which means that they take up a lot of chip area. Also, the MOST, unlike the ordinary transistor, does not have to be completely electrically insulated from the substrate, and furthermore the number of process steps is somewhat smaller. Although the miniature MOST switches are not directly accessible from outside, so that it is always necessary to have a buffer output stage which takes up a relatively large amount of space, this is not a serious drawback in large logic circuits with many internal operations or repetitive functions, such as shift registers and memories, and the internal saving in space can then be fully exploited. Depending on the size of the individual circuits, it is found in practice that MOST circuits are about one-and-a-half to three times smaller; not ten times, as is often maintained.

In itself a high packing density — often measured in terms of the number of gates per cm² — is scarcely an overriding argument for immediately choosing MOST systems. A second important factor with a considerable bearing on costs should not be overlooked, and this is the yield of the process adopted, the result of skilful helmsmanship among all manner of technological rocks and shoals, such as gaps in narrow metal tracks, the occurrence of parasitic channels or diode leakage currents, defects in epitaxial layers and faults in masks, which all too often require a great deal of patience and effort to trace, diagnose and eliminate. It will also be evident that in itself the yield is in turn a function of the desired packing density, which can make optimum costing a very complicated matter.

Finally, there is a third aspect that must not be forgotten. Once the yield of a particular process is reasonably high, the aim will then be to cut down on external connections and to integrate a number of functions; in other words, efforts will be made to combine subsystems of greater functional complexity on one chip to form a complete system. Although this involves a reduction of transit-time delays within the system, the main advantage of a minimum of connections is that of lower costs. For notwithstanding all the ingenuity and the inventiveness brought to bear on the problem of rapidly developing IC technology, we have the ironic situation that even in a large circuit the costs of encapsulation and interconnections are of the same order of magnitude as those of the inspected and tested chip. Nevertheless a limit is set to the maximum size of such an IC, because, irrespective of trivial causes of poor yield, in the present state of silicon technology the yield rapidly deteriorates when a certain chip area is exceeded. The economic limit is at present between 10 and 20 mm², but is rising only very slowly, so that for more complex integrated circuits there is more to be expected from minimizing the components than from maximizing the chip. This is one reason why the higher packing density of the MOST may sometimes tip the scales in its favour.

Apart from the density the permissible dissipation also plays a role in the MOST/bipolar controversy. Since it is almost impossible to make high-value resistors of small dimensions in conventional bipolar technology, the maximum number of transistors to be integrated is usually limited to a thousand. MOST systems involve fewer limitations, partly because of the temporary storage effect of the gate, a unique feature of the MOST enabling versions of the simple inverter to be made in which d.c. current is eliminated by the use of pulse techniques (*fig. 5*). Although this entails the use of extra circuits which regularly restore the logic information attenuated by stray leakage currents, relatively few are required in large integrated circuits. Dynamic logic circuits of this type only use energy for periodically charging or discharging minute internal capacitances, so that depending on the clock period the dissipation can be reduced by several orders of magnitude compared with static systems. If, as technology advances, more than ten thousand components can ever be accommodated in one IC, this fact will decisively tip the scales in favour of the MOST.

By about 1970 the *P*-channel MOST technique had progressed, partly because of the suitable level of the logic thresholds, to a stage that enabled a successful assault to be made on the stronghold of the bipolar transistor. In this victory the silicon gate mentioned earlier played a decisive role. One reason for this was that a low dissipation per gate could be achieved, because of the low threshold voltage and the related supply voltage, while at the same time the logic levels are well matched to those of the bipolar systems. Recently we have also seen the emergence of *N*-chan-

nel systems which, in spite of the rather more problematical threshold voltage and a greater risk of parasitic channels, are a more attractive proposition because the higher electron velocity reduces the average delay time per gate by a factor of three. Thus, while this technology leads primarily to faster systems, an alternative choice presents itself at the same time. If the greater speed is not required, there is the option of reducing the supply voltage, and this can be anticipated in the design by packing the transistors even closer together. The result is a lower energy consumption per gate and a higher packing density.

It is in fact the delay time that is the Achilles heel of MOST systems. In a circuit the delay is mainly determined by the charge that has to be fed in or out at the output by the current through one of two MOS transistors. The charge to be displaced is of course equal to the product of the output capacitance and ΔV, the difference between two logic levels. If we compare the MOST circuit with its bipolar equivalent in this respect, we find that whereas there is not much difference in the size of the displaced charges, the average charging current is so much smaller in the MOST that the final result turns out to be much less favourable for the

Fig. 4. MOST inverter, as used in integrated digital MOST circuits.

Fig. 5. Another version of the inverter circuit of fig. 4, in which there is no d.c. current because pulse techniques are used.

MOST delay time: in the simple P-channel circuit it can easily be 20 times greater. However, this very much poorer result is due not entirely to the lower effectiveness of the MOST as a switch, but partly to the higher resistance that the load MOST should have, compared with the active switch, to give an acceptable noise margin. For this reason efforts are now being made to make MOST logic a great deal faster by using various alternatives.

One of the most attractive alternatives is the complementary inverter, which can consist of an N-channel switching transistor and a P-channel load transistor. Apart from the advantage that this combination by its very nature passes no d.c. current, the switching is also faster since the output is charged or discharged by two opposing current generators.

In spite of the very much slower response of the MOST at the present time, MOST technology is already potentially capable of many and various applications in electronics. Subjects that come to mind include computer logic outside the central processor, desk calculators, electronic telephone exchanges, optoelectronics, watches and industrial control equipment.

A spectacular example of MOST circuits is to be found in the development of the semiconductor random-access memory. Here, starting with a static cell with six MOST transistors, and proceeding via a dynamic cell with three transistors, it has now been found possible to produce a dynamic memory with one MOST per cell [8]. The enormous gain in chip area achieved with this system far surpasses the extra space required for the ancillary circuits, so that a monolithic 4096 bit memory is now in an advanced stage of development. With this last example MOST technology has penetrated deeply into the world of large-scale integration (LSI), and indeed some ten thousand components can be packed in this way in an area of only 10 mm².

In this complex field MOST technology is now clearly dominant, and MOST circuits already have a 25% share of the 700 million dollar market for digital semiconductor devices. According to some estimates, this percentage will be as high as 60% in a few years time.

Should we now conclude that MOST technology is going to maintain the dominant position within the field of LSI, with perhaps a small corner reserved for the bipolar transistor in applications where speed is essential? This is very much in question, for new devel-
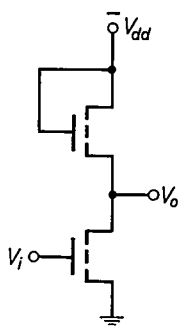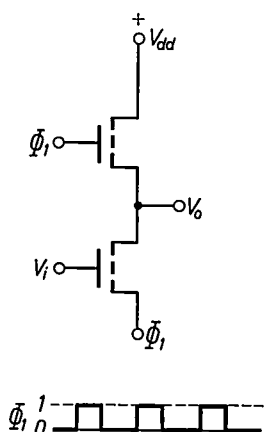
[7] See L. M. van der Steen, Digital integrated circuits with MOS transistors, Philips tech. Rev. 31, 277-285, 1970.
[8] C. W. Lambrechtse, R. H. W. Salters and L. Boonstra, A 4096-bit one-transistor per-bit RAM with internal timing and low dissipation, 1973 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 26-27.

opments are under way that could well entail a resurgence of bipolar technology. I shall touch on two of these developments here.

One line of development is connected with isolation techniques. This is understandable, because the electrical isolation required by a bipolar transistor inside that monolithic circuit is one of the features that make this transistor occupy a larger area than the MOST, which of course requires no isolation. This is often obtained by means of vertical *P-N* junctions but if they are made by the conventional diffusion methods they are rather wide. It has been found that better isolation is achieved if narrow strips of silicon oxide are used. The application of such strips is based on a local oxidation technique known as LOCOS [9] (local oxidation of silicon) or isoplanar technique. This method also offers an additional gain of space, since the relative positions of the transistors are automatically fixed and because smaller resistors can also be used. A density of fifty thousand gates is thus quite feasible, as demonstrated by a fast static 1024-bit memory brought out a few months ago, although it was rather more expensive than the dynamic MOST equivalent.

No less interesting is a second development, which is intended to correct the other shortcoming of the conventional bipolar integrated circuit: the relatively large area of high-value resistors, which are necessary to provide acceptable dissipation in large circuits. Now it was known that a complementary transistor could take over the role of a resistor, but this does not result in a more compact circuit. It is surprising that, after

some ten years of development work on digital integrated circuits, there was still room for a new elementary circuit. I refer here to the integrated injection-logic circuit (I^2L), in which a simple complementary transistor is combined with a reverse-biased bipolar switch, which can be used for making any desired logic circuit [10]. This arrangement is so compact that it can easily rival anything produced by MOST technology, although it does not at the moment look like a threat to the large MOST memories.

This concluding remark is in fact an indication that, although a fierce struggle is going on at the moment between the bipolar and MOST technologies to win the favour of the system designer, there will ultimately be room for both systems, each with their own area of applications. But there is a large area where the differences between the two are marginal, and here it will not be easy for the electronics engineer to make his choice.

[9]   See Philips tech. Rev. **31**, 276, 1970.
[10] C. M. Hart and A. Slob, Integrated Injection Logic (I^2L), Philips tech. Rev. **33**, 76-85, 1973 (No. 3).

Summary. This article is a slightly abridged version of the address given by the author in December 1972 on his inauguration as Professor Extraordinary in Semiconductor Electronics at Eindhoven University of Technology. After outlining the changes brought about in electronics by the invention of the transistor and by the advent of integrated circuits, the author turns his attention to the field-effect transistor, which had in fact long been known but has only recently become of practical interest. This applies particularly to the MOS transistor, which, in large-scale integration (LSI) has become an important competitor of the bipolar transistor. But in recent years considerable advances have also been made in logic circuits using bipolar transistors, such as injection logic (I^2L). In the long run it is to be expected that both types of transistor will find their own areas of application. Many of the differences are now marginal, and the electronics engineer is faced with a difficult choice.

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, 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 Forschungslaboratorium Aachen GmbH, Weißhausstraße, 51 Aachen, Germany   *A*

Philips Forschungslaboratorium Hamburg GmbH, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany   *H*

MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium.   *B*

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, Netherlands.

C. S. Aitchison & J. C. Williams: Actively compensated parametric amplifiers — some further results.
Electronics Letters **8**, 567-568, 1972 (No. 23).   *M*

C. Belin, J. J. Brissot & R. E. Jesse: The growth of calcite single crystals by travelling solvent zone melting.
J. Crystal Growth **13/14**, 597-600, 1972.   *L*

G. Bergmann & P. Gerthsen: (University of Karlsruhe): Electrical breakdown in gas-filled incandescent lamps.
Lighting Res. Technol. **4**, 159-165, 1972 (No. 3).   *A*

F. Berz: Fluctuations in noise level.
Electronics Letters **8**, 515-517, 1972 (No. 21).   *M*

G. Bioul, M. Davio & J. P. Deschamps: Minimization of ring-sum expansions of Boolean functions.
Philips Res. Repts. **28**, 17-36, 1973 (No. 1).   *B*

D. Boccon-Gibod & J.-L. Teszner: (Ecole Normale Supérieure, Paris): Effets bidimensionnels et de surface dans une diode Gunn à contacts coplanaires.
Acta Electronica **15**, 205-216, 1972 (No. 3).   *L*

L. Boonstra & F. L. J. Sangster: Progress on bucket-brigade charge-transfer devices.
1972 IEEE Int. Solid-State Circuits Conf. Digest tech. Papers, pp. 140-141 & 228.   *E*

P. W. J. M. Boumans & F. J. de Boer: Studies of flame and plasma torch emission for simultaneous multi-element analysis, I. Preliminary investigations.
Spectrochim. Acta **27B**, 391-414, 1972 (No. 9).   *E*

J. P. Boutot & G. Piétri: Photomultiplier control by a clamping cross-bar grid.
IEEE Trans. **NS-19**, No. 3, 101-106, 1972.   *L*

P. Branquart: Contribution to symposium: ALGOL 68 implementation.
Information Processing 71, Proc. IFIP Congress, Ljubljana 1971, Vol. 2, pp. 1580-1582, 1972.   *B*

P. Branquart, J. P. Cardinael, J. P. Delescaille & J. Lewi: A context-free syntax of ALGOL 68.
Inform. Proc. Letters **1**, 141-148, 1972 (No. 4).   *B*

D. J. Breed & R. P. Kramer: Charge trapping in MOS systems.
Thin Solid Films **13**, 1-4, 1972 (No. 1).   *E*

D. J. Breed & R. P. Kramer: A new type of charge trapping in MOS systems.
Philips Res. Repts. **28**, 75-79, 1973 (No. 1).   *E*

R. Brun, J. Cayzac & R. Genève: Problèmes relatifs aux liaisons mobiles pour téléreportage couleur.
Rev. techn. Radiodiff. Télév. No. 22, 25-35, 1972.   *L*

T. M. Bruton & E. A. D. White: (Imperial College of Science and Technology, London): Low frequency linear electro-optic effect in $PbTa_2O_6$.
J. Mat. Sci. **7**, 1233-1234, 1972 (No. 11).   *M*

K. Bulthuis: Time history of laser power pulses from molecular gas lasers.
Physics Letters **42A**, 38-40, 1972 (No. 1).   *E*

K. H. J. Buschow & J. H. N. Creyghton: Magnetic properties of rare earth tetraborides.
J. chem. Phys. **57**, 3910-3914, 1972 (No. 9).   *E*

K. L. Bye & E. T. Keve: Structural inhibition of ferroelectric switching in triglycine sulphate, II. X-irradiation/field treatment.
Ferroelectrics **4**, 87-95, 1972 (No. 2).   *M*

F. M. A. Carpay: Some remarks on the λ-R relationship in eutectoid decompositions.
Scripta metall. **6**, 1019-1022, 1972 (No. 10).   *E*

H. B. G. Casimir: Impact of computers on industrial research and development.
Computer Phys. Comm. 3, Suppl., 174-178, 1972 .   *E*

J. P. Chané, L. Hollan & C. Schiller: Etude de l'épitaxie localisée du GaAs.
J. Crystal Growth 13/14, 325-330, 1972.   L

J. B. Clegg, I. G. Gale & E. J. Millett: Determination of oxygen in semiconductor materials with a cryogenically pumped spark-source mass spectrometer.
Analyst 98, 69-74, 1973 (No. 1162).   M

L. J. Collins, R. F. Hall, K. E. Johnson & G. T. Sharpless: A 56-character data display using a d.c. glow discharge matrix panel.
SERT J. 7, 7-10, 1973 (No. 1).   M

J. P. M. Damen & J. M. Robertson: Induced nonperiodic growth striations in flux-grown magnetic oxide single crystals.
J. Crystal Growth 16, 50-53, 1972 (No. 1).   E

N. Dedding (Philips Radio, Television and Record-playing Equipment Division, Eindhoven) & J. A. W. van Laar (Utrecht State University): Selection criterion for a reflectometer.
Verfkroniek 45, 311-316, 1972 (No. 11).

Ph. Delsarte: Weights of linear codes and strongly regular normed spaces.
Discrete Math. 3, 47-64, 1972 (No. 1/2/3).   B

R. Dessert, J. Magarshack & V. Pauker (RTC La Radiotechnique-Compelec, Suresnes, France): Applications des diodes à effet Gunn.
Acta Electronica 15, 191-204, 1972 (No. 3).   L

P. A. Devijver: On an asymptotic property of the least-mean-square-error design criterion in pattern recognition.
Philips Res. Repts. 28, 37-49, 1973 (No. 1).   B

H. Dimigen, H. Hieber, H. Hoffmann, H. W. Neuhaus, L. Stewen & J. Verweel: Closed-flux elements for integrated magnetic memories.
IEEE Trans. MAG-8, 597-599, 1972 (No. 3).   H

H. Dimigen & J. Verweel: Some technological aspects of integrated closed-flux magnetic memory elements.
Int. J. Magnetism 3, 197-205, 1972 (No. 1-3).   H

J. A. W. van der Does de Bye, A. C. P. van den Bosch, C. M. Hart, M. Saitoh, A. Slob & P. Verhoog: Multichannel photon counter for luminescence decay measurements.
Rev. sci. Instr. 43, 1468-1474, 1972 (No. 10).   E

W. G. Essers, G. Jelmorini & G. W. Tichelaar: Arc characteristics and metal transfer with plasma-MIG welding.
Metal Constr. Brit. Welding J. 4, 439-447, 1972 (No. 12).   E

E. Fischer, R. Lorenz & L. Rehder: Lichterzeugung durch quasikontinuierliche Molekularstrahlung in Hochdruck-Gasentladungslampen.
Lichttechnik 24, 513-516, 1972 (No. 10).   A

A. T. Florence (R. J. Reynolds Industries Inc., Winston-Salem, N.C., U.S.A.) & G. Frens: Aureole profile in bursting soap films. Surface tension and surface relaxation in rapidly compressed monolayers.
J. phys. Chem. 76, 3024-3029, 1972 (No. 21).   E

K. G. Freeman & D. B. Spencer: Experimental receivers for 12 GHz f.m. television from satellites.
IEE Conf. Publn. No. 88: International Broadcasting Convention, pp. 33-39, 1972.   M

G. Frens: Particle size and sol stability in metal colloids.
Kolloid-Z. & Z. Polymere 250, 736-741, 1972 (No. 7).   E

M. Gautherie (Centre Hospitalier Universitaire de Strasbourg), M. Jatteau, J. Ott & Y. Quenneville (Centre Hosp. Univ. Strasbourg): Mesure précise de températures cutanées par thermographie infrarouge.
Path. & Biol. 20, 559-575, 1972 (No. 11-12/13-14).   L

J. A. Geurst & W. J. A. Goossens: Theory of electrically induced hydrodynamic instabilities in smectic liquid crystals.
Physics Letters 41A, 369-370. 1972 (No. 4).   E

J. M. Goethals & S. L. Snover (Eindhoven University of Technology): Nearly perfect binary codes.
Discrete Math. 3, 65-88, 1972 (No. 1/2/3).   B

W. J. A. Goossens: The conduction regime in a nematic liquid crystal with negative dielectric anisotropy at frequencies above the dielectric relaxation frequency.
Physics Letters 40A, 95-96, 1972 (No. 2).   E

G. G. P. van Gorkom: Theoretical optical Zeeman patterns for Mn^{2+} in an octahedral or tetrahedral crystal field.
Solid State Comm. 11, 1253-1256, 1972 (No. 9).   E

D. Gossel: Frequenzanalogie. Ein Konzept für Meß- und Regelsysteme mit digitaler Signalverarbeitung.
Elektrotechn. Z. A 93, 577-581, 1972 (No. 10).   H

M. Grenot, J. Pergrale, J. Donjon & G. Marie: New electro-optic light valve device for image storage and processing.
Appl. Phys. Letters 21, 83-85, 1972 (No. 3).   L

G. Groh: Some optical information processing systems.
Optical and acoustical holography, editor E. Camatini, publ. Plenum Press, New York 1972, pp. 71-110.   H

J. de Groot & A. Mircea: Diodes Gunn: principes physiques et calculs de simulation.
Acta Electronica 15, 145-161, 1972 (No. 3).   E, L

P. F. A. Haans & P. Heller: Een methode voor het aansmelten van ronde glasplaatjes op glazen buizen.
Glastechn. Meded. 9, 167, 1971 (No. 5).   E

P. F. A. Haans & P. Heller: Methode voor het maken van flenzen aan glazen cilinders en buizen.
Glastechn. Meded. 9, 170-171, 1971 (No. 5).   E

P. F. A. Haans & H. F. Wilbrink: Het maken van gaten in glas.
Glastechn. Meded. 9, 125-127, 1971 (No. 4).   E

H. B. Haanstra, W. F. Knippenberg & G. Verspui: Columnar growth of carbon.
J. Crystal Growth 16, 71-79, 1972 (No. 1).   E

W. van Haeringen, M. F. H. Schuurmans & H.-G. Junginger: Analysis of electronic structure using empirical pseudo-potentials.
J. Physique 33, C3/185-190, 1972 (Colloque C3).   E, A

H. W. Hanneman & H. N. Linssen (Eindhoven University of Technology): A relation between sensitivity and variance of network functions.
IEEE Trans. **CT-19**, 499-502, 1972 (No. 5).   *E*

P. Hansen & W. Tolksdorf: Ferromagnetic resonance in Al^{3+}-substituted yttrium iron garnet with small amounts of Ru^{3+}.
Int. J. Magnetism **3**, 81-85, 1972 (No. 1-3).   *H*

P. Hansen, W. Tolksdorf & J. Schuldt: Anisotropy and magnetostriction of germanium-substituted yttrium iron garnet.
J. appl. Phys. **43**, 4740-4746, 1972 (No. 11).   *H*

K. Hart & A. Slob: Integrated injection logic: a new approach to LSI.
IEEE J. **SC-7**, 346-351, 1972 (No. 5).   *E*

H. Haug & K. Weiss: Reduction of the Kapitza resistance through surface dislocation scattering.
Proc. Int. Conf. on Phonon Scattering in Solids, Paris 1972, pp. 386-388.   *E*

E. E. Havinga: Oscillatory dependence of electronic properties of alloys with Cu_3Au-type structure; band structure calculations.
Solid State Comm. **11**, 1249-1252, 1972 (No. 9).   *E*

B. Hill: Point efficiency and signal-to-background ratio in exponential holograms for optical memories.
Appl. Optics **11**, 2937-2944, 1972 (No. 12).   *H*

B. Hoekstra, R. P. van Stapele & A. B. Voermans: Magnetic anisotropy of tetrahedral ferrous ions in $CdCr_2S_4$.
Phys. Rev. B **6**, 2762-2769, 1972 (No. 7).   *E*

K. Holford: Doppler radar with sense.
Wireless World **78**, 535-539, 1972 (No. 1445).   *M*

L. Hollan & C. Schiller: Etude de l'anisotropie de la croissance épitaxiale de GaAs en phase vapeur.
J. Crystal Growth **13/14**, 319-324, 1972.   *L*

A. P. Hulst: Macrosonics in industry, 2. Ultrasonic welding of metals.
Ultrasonics **10**, 252-261, 1972 (No. 6).   *E*

H.-G. Junginger, W. van Haeringen & M. F. H. Schuurmans: Computation of electronic properties of solids.
Computer Phys. Comm. **3**, Suppl., 127-136, 1972. *A, E*

D. Kasperkovitz: A 70 MHz static shift register with high integration density.
Microelectronics and Reliability **11**, 493-497, 1972 (No. 6).   *E*

D. Kasperkovitz: An all-bipolar image sensor.
Solid-State Electronics **16**, 67-73, 1973 (No. 1).   *E*

J. T. Klomp: Bonding of metals to ceramics and glasses.
Amer. Ceramic Soc. Bull. **51**, 683-688, 1972 (No. 9).   *E*

J.-P. Krumme & J. Haberkamp: Critical assessment and modification of Robinson's method for the ellipsometric investigation of thin film materials.
Thin Solid Films **13**, 335-339, 1972 (No. 2).   *H*

K. Lagemann: Das Blockschachtelungsverfahren, eine Strategie beim rechnergestützten Entwurf von integrierten Großschaltungen (LSI).
Elektron. Rechenanl. **14**, 202-205, 1972 (No. 5).   *H*

G. Le Floch: Modification of the *c*-axis orientation of hexagonal cadmium sulfide films by a new technique of evaporation by reflection.
Solid-State Electronics **15**, 753-756, 1972 (No. 7).   *L*

J. Magarshack, A. Mircea & A. Roussel: Bruit BF dans GaAs de type *n* et sa corrélation avec le bruit HF des oscillateurs à diodes Gunn.
Acta Electronica **15**, 233-243, 1972 (No. 3).   *L*

R. Memming: Photochemical and electrochemical processes of excited dyes at semiconductor and metal electrodes.
Photochem. Photobiol. **16**, 325-333, 1972 (No. 4).   *H*

R. Metselaar & P. Rem: Radio frequency sputtering of ferrites.
Proc. 1st European Sputtering Symp., Toulouse 1969, pp. 58-63; 1972.   *E*

F. Meyer & J. J. Vrakking: Quantitative aspects of Auger electron spectroscopy.
Surface Sci. **33**, 271-294, 1972 (No. 2).   *E*

B. J. Mulder: Recombination diffusion length of minority charge carriers in cuprous sulphide (chalcosite and djurleite).
Phys. Stat. sol. (a) **13**, 569-575, 1972 (No. 2).   *E*

H. W. Newkirk: Piezoelectric organometallic crystals: growth and properties of tetraphenylsilicon, -germanium, -tin and -lead.
J. organomet. Chem. **44**, 263-271, 1972 (No. 2).   *A*

D. H. Paxman: Indium phosphide as a new material for microwave (transferred electron effect) oscillators.
Acta Electronica **15**, 245-252, 1972 (No. 3).   *M*

P. Piret: A stochastic model for burst-correcting convolutional decoders.
IEEE Trans. **IT-18**, 532-535, 1972 (No. 4).   *B*

A. Rabier: Amplificateurs à diodes Gunn.
Acta Electronica **15**, 183-190, 1972 (No. 3).   *L*

O. Reifenschweiler: The generation of high neutron yields with sealed-off neutron tubes.
Proc. 1st Symp. on Neutron Dosimetry in Biology and Medicine, Neuherberg/München 1972, pp. 821-837. *E*

F. Rondelez & J. P. Hulin: Distortions of a planar cholesteric structure induced by a magnetic field.
Solid State Comm. **10**, 1009-1012, 1972 (No. 11).   *L*

B. Schiek & K. Schünemann (Valvo, Hamburg): Noise of negative resistance oscillators at high modulation frequencies.
IEEE Trans. **MTT-20**, 635-641, 1972 (No. 10).   *H*

W. Schilz: Spin-wave propagation in epitaxial YIG films.
Philips Res. Repts. **28**, 50-65, 1973 (No. 1).   *H*

P. J. Severin & G. J. Poodt: Capacitance-voltage measurements with a mercury-silicon diode.
J. Electrochem. Soc. **119**, 1384-1389, 1972 (No. 10). *E*

L. A. Æ. Sluyterman & J. Wijdenes: Sigmoidal progress curves in the polymerization of leucine methyl ester catalysed by papain.
Biochim. biophys. Acta **289**, 194-202, 1972 (No. 1).   *E*

R. Spitalnik: Oscillations à croissance de charge d'espace contrôlée dans l'arséniure de gallium.
Acta Electronica **15**, 163-181, 1972 (No. 3).   *L*

P. J. A. Strik & H. F. Wilbrink: Schroefdraad op glasbuis.
Glastechn. Meded. **9**, 158-166, 1971 (No. 5).   *E*

A. Thayse: Disjunctive and conjunctive operators for Boolean functions.
Philips Res. Repts. **28**, 1-16, 1973 (No. 1).   *B*

R. Veilex: Introduction *(to second Gunn-effect issue of Acta Electronica).*
Acta Electronica **15**, 141-142 (in French), 143-144 (in English), 1972 (No. 3).   *L*

J. F. Verwey: Avalanche-injected hole current in SiO_2.
Appl. Phys. Letters **21**, 417-419, 1972 (No. 9).   *E*

J. F. Verwey: Time dependence of injection currents in SiO_2.
Philips Res. Repts. **28**, 66-74, 1973 (No. 1).   *E*

A. T. Vink & G. G. P. van Gorkom: Characteristic infrared luminescence in GaP due to Mn.
J. Luminescence **5**, 379-384, 1972 (No. 5).   *E*

W. L. Wanmaker & J. G. Verriet (Philips Lighting Division, Eindhoven): Luminescence of phosphors with $Ca_3SiO_4Cl_2$ as the host lattice.
Philips Res. Repts. **28**, 80-83, 1973 (No. 1).

K. Weiss: Eine elektrochemische Methode zur Abschätzung der Oberflächenenergie von α-Ag_2S.
J. Solid State Chem. **5**, 398-401, 1972 (No. 3).   *E*

J. Wolter: Controllable parametric amplification of Bleustein-Gulyaev plate waves in Si-PXE 5 structures.
Physics Letters **42A**, 115-116, 1972 (No. 2).   *E*

H. Zijlstra: De coërcitiefkracht van permanente magneten.
Ned. T. Natuurk. **38**, 325-331, 1972 (No. 19).   *E*

Contents of Philips Telecommunication Review **31**, No. 1, 1973:

Contents of Mullard Technical Communications **12**, No. 117, 1973:

Contents of Mullard Technical Communications **12**, No. 118, 1973:

Contents of Valvo Berichte **17**, No. 2, 1973:

Small electric motors

The history of the electric motor is already a long one. A recent feature is the marked increase, in the last ten years or so, in the use of electric motors with a power rating of less than one kilowatt. Nowadays hundreds of millions of these small motors are manufactured every year, and in some countries the total value of the annual production exceeds that of the larger motors.

There are two main destinations for all these small motors: domestic use and automation. In the modern home they perform many duties — one motor may provide a mechanical power of less than one watt for a record player, another a power of several hundred watts for a large domestic appliance — and today a total of twenty or more small electric motors can easily be found in a single home. In the automation of industrial equipment small electric motors are used for all kinds of control, adjustment and other processes that were formerly carried out by hand.

Since Philips are manufacturers of consumer articles and industrial control equipment they have become manufacturers of large numbers of small electric motors; the annual production now reaches tens of millions. This means that through the years a great deal of research and development work has been done on motors; yet very little on the subject has appeared in our journal.

In the present issue and in two later ones, which will appear in the next eighteen months, we wish to bring the motor research at Philips somewhat more into the limelight. We have selected a number of very different subjects from this research directed towards small motors — often extremely simple in construction to keep the costs down, but with the theory just as elaborate — and each of our 'motor' issues will therefore have a very varied content. Thus articles will be found on progress in the theoretical fields (single-phase synchronous motor, battery motor), as well as descriptions of new constructions (vibrator, moving-coil motor, brushless d.c. motor) and methods for electronic control of the speed of d.c. and induction motors. There are also a few articles of a more general nature. One of these appears as the introduction at the front of this first motor issue; it describes the operation and the characteristics of the various existing types of electric motor, and can be taken as a fundamental background to the other articles.

Electric motors

E. M. H. Kamerbeek

Types of electric motor

The function of an electric motor is to convert electrical energy into mechanical work. The electrical energy supplied is usually obtained from mechanical work applied simultaneously somewhere else in a conversion in the other direction by means of generators. Electric motors and generators are both referred to as electrical machines and are essentially the same in construction; in principle the same electrical machine can perform both functions.

In this article a description will be given of the fundamental principles of the operation of the electrical machine. Various types of machine will be considered, all of them types that can be applied as small electric motors. Here small motors are taken to be motors with a rating of 1 kilowatt or less.

Most electric motors carry out the energy conversion by way of a rotating movement — exceptions are the linear electric motor, in which the movement is a linear one, and the vibrator [1]. The rotating part of the machine is called the rotor, and the stationary part the stator. At least one of these is provided with one or more windings that are connected to the source of electric power. Depending on the type of source, the machine will be a *d.c. motor* or an *a.c. motor*.

In both cases the operation of the motor can often be described by starting with the Lorentz force — the force experienced by a conductor carrying a current in a magnetic field (see *fig. 1*). If we imagine that the magnetic field in a d.c. motor is stationary, then to ensure that the Lorentz force on a conductor at the circumference of the rotor has the correct direction during the whole of each revolution, the current must be regularly reversed. This is the function of the commutator, an essential feature of every d.c. motor [2]. The commutator is usually a mechanical device and generally consists of a cylinder formed by insulated conducting segments mounted on the shaft; the windings are connected to the segments and the current is fed to the segments by brushes (*fig. 2*).

A commutator of this type is subject to wear, particularly when there is sparking between brushes and segments. The sparking also causes radio-frequency interference. For these reasons there has been consider-

able interest recently in developing electronic methods of commutation; motors with an 'electronic commutator' are also called brushless d.c. motors.

The d.c. motor is the obvious choice when battery supplies have to be used. A characteristic feature of the d.c. motor is that its speed of rotation can be varied over a wide range. When this is a requirement, d.c. motors are sometimes used even though power has to be taken from a.c. mains [3]. The current then has to be rectified; modern circuits for rectification and control include semiconductor diodes, transistors and thyristors.
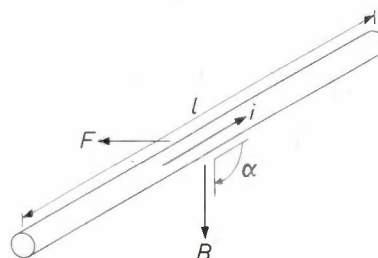
Fig. 1. The Lorentz force. A conductor carrying a current *i* in a magnetic field of flux density *B* is subject to a force *F*, the Lorentz force. In the diagram the conductor is assumed to be straight. *F* is at right angles to the plane of the vectors *i* and *B*; the magnitude *F* of the force is proportional to the sine of the angle α between *i* and *B*, the length *l* of the conductor, the strength of the current and the magnetic flux density. In vector notation $F = li \times B$. In motors *i* and *B* are usually perpendicular so that the magnitude of the force is given by $F = liB$.
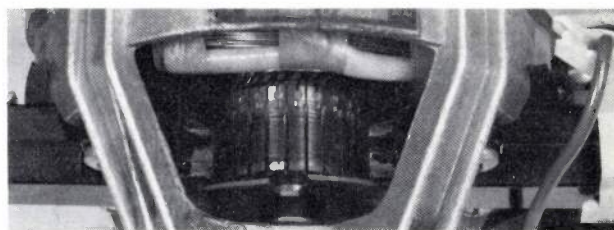
Fig. 2. Commutator. Each segment is connected to a part of the rotor winding. The current is supplied via two stationary brushes, which make a sliding contact with the segments.

[1] J. Timmerman, Two electromagnetic vibrators; this issue, p. 249.

[2] An exception is the unipolar machine, but this is rarely used. A special form of this machine, which made use of superconductivity, has been described by J. Volger in: A dynamo for generating a persistent current in a superconducting circuit, Philips tech. Rev. **25**, 16-19, 1963/64.
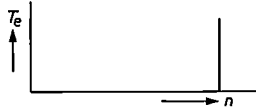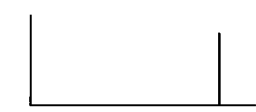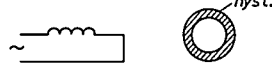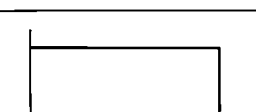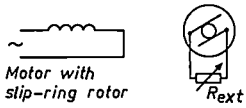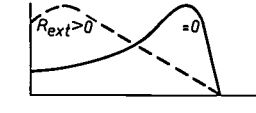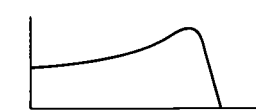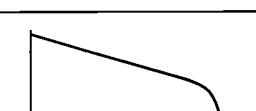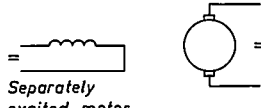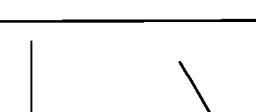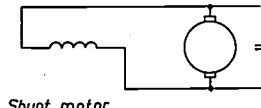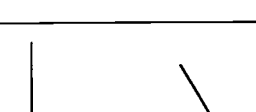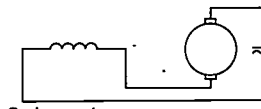
[3] See the article by R. Raes and J. Schellekens about a d.c. motor with speed control for a washing machine, to appear in the second issue on electric motors.

Dr Ir E. M. H. Kamerbeek is with Philips Research Laboratories, Eindhoven.

In a.c. machines there is no need for a commutator; use can be made here of the periodic change in the direction of the current in the a.c. mains. This means that a relation must exist between the frequency of the a.c. current and the speed of revolution of the machine. In *synchronous* machines the number of revolutions per second is proportional to the frequency of the current. Synchronous machines are therefore widely applied where constant speed is a first requirement. Then there is the group of *asynchronous* motors; these motors run at a speed a little lower than the 'synchronous' value, and the difference (the slip speed) increases with the

load on the motor. Because of this difference in speed currents are induced in closed circuits in the rotor; this induction is essential to the operation of this type, which is therefore usually referred to as the *induction motor*. The induction motor is the most common type of a.c. motor.

In some applications the relation between the speed and the mains frequency is a nuisance. One way around the difficulty is to use a commutator, even though the supply is a.c. This is because a commutator motor in the correct kind of circuit can operate from a mains supply of any frequency — and not just if that fre-

Table I. Survey of the most important types of motor. The torque-speed characteristics show the variation of the electromagnetic torque T_e as a function of the rotational speed n of the motor.

	Type	Torque-speed characteristic	Special features	See page
Synchronous motors	Motor with permanent-magnet rotor		Also applied as stepping motor and brushless d.c. motor	217
	Reluctance motor		Also applied as stepping motor and brushless d.c. motor	223
	Hysteresis motor		Self-starting	224
Asynchronous or induction motors	Motor with slip-ring rotor		An external resistance R_{ext} in the rotor circuit can be used for starting	226
	Squirrel-cage motor		Torque-speed characteristic affected by shape of cage (double cage, skin-effect cage)	
	Motor with solid rotor		Skin effect in the rotor is of major importance	
Commutator motors	Separately excited motor		Permanent magnet often used in small motors instead of field winding	230
	Shunt motor			
	Series motor		Can also be used for a.c. (universal motor)	

quency is zero, as for a d.c. motor. In practice a motor to be used in this way has to be specially designed, and is called a universal motor. If such a motor is run from an a.c. supply the currents in the rotor and stator windings continually change sign simultaneously. This means that the interactive force, which is proportional to the product of the two currents, maintains the same sign.

The speed of a synchronous or asynchronous motor can also be varied over a wide range if it is supplied from a source of variable frequency. This requires the mains frequency to be converted to another frequency; there are rotary frequency converters (a motor driving a generator) and static (or electronic) frequency converters. In fact the operation of the commutator in a d.c. or universal motor can also be considered as a frequency conversion. This is why the combination of a synchronous motor and a frequency converter is found to behave in a similar way to a d.c. commutator motor under certain conditions.

Finally, there is the stepping motor [*]. This is essentially a motor of the synchronous type, supplied by electronically generated pulses of current. Each pulse produces a constant angular rotation. The number of angular rotations (the 'steps') is equal to the number of pulses applied, so that a digital signal is converted into a mechanical displacement. Stepping motors are used in applications such as numerically controlled machine tools.

Motors of the various basic types listed here, except perhaps the stepping motor, were in existence long be-fore the end of the nineteenth century [**]. Nevertheless, new developments are still under way. The main trends are:
— Improvement of the various characteristic features (the efficiency, the variation of torque with speed, the starting characteristic, etc.).
— Increasing the power/weight and power/volume ratios.
— Improving reliability and safety.
— Improving methods of motor control.
These developments are possible since improved magnetic and insulating materials are now available, computers can be used in design work, and new production techniques can be used, such as sintering of rotors. The availability of electronic components, in particular thyristors and transistors, has also increased the range of possibilities, especially in motor control.

We shall now examine the operation and characteristics of the most important types of motor in more detail, and derive an expression for the delivered torque. Table I shows a diagrammatic survey of the motor types examined in this article, with the page references. Types with an external rotor, sometimes used in small motors, are not described separately as they do not differ in principle from the types with an internal rotor. Since the synchronous motors to be discussed in this article are of simple configuration, we shall start with these. Asynchronous or induction motors and the commutator motors will then follow.

I. Synchronous motors

The kinds of synchronous motor to be discussed here all have a rotor with no windings. This rotor can be a permanent magnet, in which case its magnetization will give it a preferential position with respect to the magnetic field produced by the stator windings. In another type, the reluctance motor, the rotor is not permanently magnetized, but is made from a material of high permeability; this rotor is not cylindrical and therefore has a preferential position for which the magnetic resistance — the reluctance — of the magnetic field is at a minimum. In both motors the stator windings excite a rotating magnetic field; the rotor tries to maintain the preferential position with respect to the field and rotates with it, provided that it is not prevented from doing so by too great a load on the shaft. In the third type of motor to be discussed, the hysteresis motor, there is also a rotating stator field. This induces a magnetization in the rotor, as in the reluctance motor. The rotor is now cylindrical, however, and made of a material that has considerable magnetic hysteresis. Be-cause of this hysteresis the magnetization always lags behind the direction of the rotating stator field, thus giving a torque.

Synchronous motor with permanent-magnet rotor and cylindrical stator bore

Let us now pass on to a calculation of the torque operating on the rotor of a synchronous motor with a permanent-magnet rotor. We shall assume here that the motor is constructed as in the cross-sectional diagram of *fig. 3*. A cylindrical rotor can rotate in the cylindrical bore of the stator. The stator — usually laminated in practice to reduce eddy currents, i.e. made from a stack of thin stampings insulated from each other — carries two 'diametral' windings *1, 1'* and *2,2'*. In fig. 3 these are shown in the air gap, but in practice

[*] We intend to publish a separate article on this subject in the third issue on electric motors. *(Ed.)*
[**] We intend to include an article on the history of hte electric motor in the third issue on electric motors. *(Ed.)*

they can be mounted in slots in the stator. The rotor is homogeneously magnetized parallel to a diameter.

It is not difficult to calculate the forces experienced by the sides of the coils in the magnetic field of the rotor when a current flows. These forces taken together give a torque on each coil. The reaction torque on the rotor is then equal in magnitude to the sum of the torques on the two coils but of opposite sense. The only component of the magnetic field that contributes to the torque is the radial component that crosses the air gap. This component is therefore the only one considered in the rest of the discussion. Similarly, the forces experienced by the coils from their own and one another's field will not be taken into account, since they do not contribute to the torque.
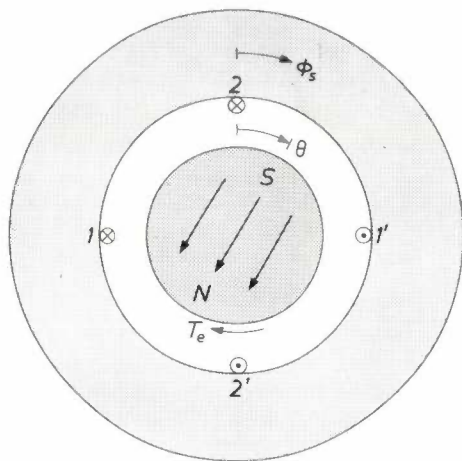
Fig. 3. Schematic cross-section of a synchronous motor with a cylindrical permanent-magnet rotor (N north pole, S south pole). The rotor rotates inside the cylindrical bore of a stator carrying two 'diametral' coils $1,1'$ and $2,2'$ When current flows in the coils in the magnetic field of the rotor they experience a torque; a reaction torque T_e of the same magnitude but opposite sign then operates on the rotor. If suitable a.c. currents flow in the coils and the rotor has the correct angular velocity, the torque on the rotor then always has the same sign and the rotation is maintained. The position of the rotor is given by the angle θ; a rotation in the direction of increasing θ is taken as positive. The angle ϕ is the coordinate used in the calculation for position on the stator circumference.

If a current $I_{s(1)}$ flows in the coil 1, then a side of this coil of length l (at right angles to the plane of the drawing in fig. 3) will experience a Lorentz force $F = I_{s(1)}N_s lB$, where B is the magnetic flux density in the air gap and N_s is the number of turns in the coil. Using B everywhere for the radial component, as indicated above, then whatever the position of the rotor F is always perpendicular to the plane of the coil. The magnitude of B varies sinusoidally along the circumference of the rotor: $B = \hat{B} \cos \phi_r$, where ϕ_r is the angle in radians with respect to the plane of symmetry of the rotor. For an arbitrary angle θ of the rotor with respect to the stator the variation of B around the stator

is given by $B = \hat{B} \cos (\phi_s - \theta)$; here ϕ_s is the coordinate along the circumference of the stator (see fig. 3). The force on the side of the winding at $\phi_s = \pi/2$ is then $F = I_{s(1)}N_s l\hat{B} \sin \theta$; the force on the other side (at $\phi_s = -\pi/2$) is equal in magnitude but of opposite sense. The torque resulting from the two forces is $2aF$, where $2a$ is the width across the coil. If the second winding shown in fig. 3 is not taken into account this is the electromagnetic torque acting in the motor. With the directions of current and magnetization of the rotor as shown in fig. 3 the sense of the reaction torque on the rotor is opposite to the direction of increasing θ, i.e. it is negative, so that for this torque we have:

$$T_{e(1)} = -2alI_{s(1)}N_s\hat{B} \sin \theta. \tag{1}$$

The torque is zero in the two positions opposite each other for which θ is equal to zero or π radians. In the first position the equilibrium is stable; for a small deviation of the rotor it experiences a torque tending to return it to this position. This is the preferential position referred to above. In the second position the equilibrium is unstable; the sense of the torque is such that a small deviation will increase.

By making use of the coil 2, which is rotated by $\pi/2$ radians with respect to the first coil, the rotor can be given another stable final position $\pi/2$ radians ahead of or behind the first one. In this way the rotor can be kept rotating in 'steps'. This is the essential principle of a stepping motor with a permanent-magnet rotor and a 'stepping angle' of $\pi/2$ radians.

However, what is usually required is a steady continuous rotation of the rotor. This can be obtained by supplying the coils 1 and 2 with sinusoidally varying a.c. currents of equal magnitude and differing in phase by $\pi/2$ radians. To calculate the torque in this case the expressions for the two currents, $i_{s(1)} = \hat{i}_s \cos \omega t$ and $i_{s(2)} = \hat{i}_s \sin \omega t$ (where ω is the angular frequency of the a.c. currents and t the time) can be substituted in equation (1). Summing the torques produced by the two coils will give the total torque T_e. If it is now further assumed that the rotor rotates at a constant angular velocity ω_r, so that $\theta = \omega_r t + \theta_0$, then for the total torque we have:

$$T_e = -2al\hat{i}_s N_s\hat{B} \sin\{(\omega_r - \omega)t + \theta_0\}. \tag{2}$$

The time-average value of this torque can only differ from zero if ω_r is equal to ω, i.e. if the rotor runs at synchronous speed. This means that the motor will not operate as a motor at any speed other than the synchronous speed. It also means that the motor will not start of its own accord, since the time average of the torque at zero speed is also zero. This is an important practical difficulty; to avoid it synchronous motors are often provided with a second rotor of another type on

the same shaft and rotating in the same stator. The second rotor can be for example a squirrel-cage rotor, as used in induction motors (see below).

At the synchronous speed the torque is constant:

$$T_e = -2al\hat{i}_s N_s \hat{B} \sin \theta_0. \tag{3}$$

With the motor rotating at constant speed, the accelerating torque acting on the rotor is apparently zero. This means that the magnitude of the electromagnetic torque is exactly large enough to compensate the torques produced by the load and by friction. At a given current the only variable in (3) that could enable the motor to match the electromagnetic torque to the load is the angle θ_0 between the rotor and the stator field. For motor operation θ_0 is negative; the rotor lags behind the rotating stator field, falling further back as the load increases. When θ_0 is equal to $-\pi/2$ the electromagnetic torque has reached its maximum; if the load increases further the motor will be unable to maintain its constant speed of rotation; it will 'fall out of step', and come to rest. This shows the limited flexibility of the synchronous motor, which cannot maintain the equilibrium between the electromagnetic and mechanical torque by a reduction in speed.

The synchronous motor will operate as a generator when the rotor is externally driven. The effect of the electromagnetic forces on the rotor is then to produce a braking torque, of the opposite sense to that in motor operation. In this arrangement θ_0 in equation (3) is now positive.

Higher harmonics

So far it has been assumed that the magnetic field originating from the rotor is sinusoidally distributed along its circumference. This is usually only approximately true. Terms of higher order are therefore present; these give rise to fluctuations in the torque at every revolution of the rotor, which interfere with the even running of the motor.

These fluctuations can be counteracted by using distributed windings *1* and *2* instead of the diametral coils *1* and *2*. A distributed winding consists of a number of coils at various points of the stator circumference, all connected in series. In this case it is more convenient to replace the quantity $N_s i_{s(1)}$ by the current-density distribution $S_{s(1)}(\phi_s)$, expressed in amperes per unit of length along the stator circumference. Since all the turns are in series, $S_{s(1)}$ is equal to $i_{s(1)} z_s(\phi_s)$, where z_s is the number of turns per unit of length. This quantity z_s is called the copper-density distribution. The sign of this function indicates whether the current at the location ϕ_s flows in the positive or negative direction ($\otimes$ or $\bigcirc$ in fig. 3) when a positive current is applied at the terminals.

In many cases a distributed winding in which $S_{s(1)}$ is distributed sinusoidally around the stator circumference is ideal. As we saw above, a uniform torque can be obtained with a purely sinusoidal distribution of the rotor flux density and a non-sinusoidal current-density distribution. An even torque is also obtained with a sinusoidal current-density distribution and a magnetic flux density that is not sinusoidally distributed, since the current i and the magnetic flux density B are equivalent factors in the expression $F = liB$ for the Lorentz force.

In practice only an approximation to a sinusoidal current-density distribution is attempted. The coils of the winding are usually arranged in equidistant slots in the stator circumference. *Fig. 4* shows a stator with
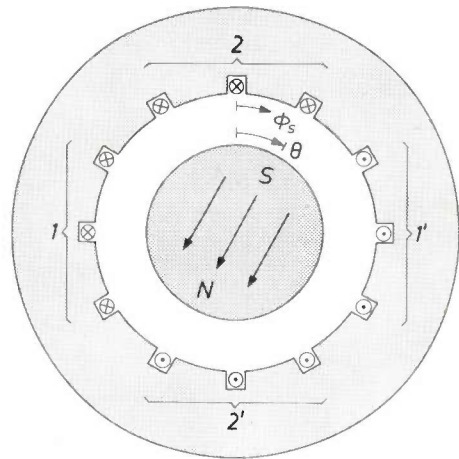
Fig. 4. Schematic cross-section of a synchronous motor in which the two windings *1,1'* and *2,2'* each consist of three coils. The stator has twelve equidistant slots for the coils. This arrangement is used to give an approximation to a sinusoidal current-density distribution.

twelve slots, in which there are two windings, each consisting of three coils. Usually two to five coils per winding are considered sufficient, with equal numbers of turns in each coil. This approach gives a considerable reduction in the torque fluctuation.

If the distribution of the rotor flux density B_r around the stator circumference is symmetrical but not sinusoidal, it will contain higher harmonics and we can write:

$$B_r(\phi_s) = B_{r1} \cos(\phi_s - \theta) + B_{r2} \cos 2(\phi_s - \theta) + \\ + B_{r3} \cos 3(\phi_s - \theta) + \dots .$$

At the synchronous angular velocity ω the torque is then

$$T_e = -2al\hat{i}_s N_s \{ B_{r1} \sin \theta_0 - B_{r3} \sin(4\omega t + 3\theta_0) + \\ + B_{r5} \sin(4\omega t + 5\theta_0) - \dots \}.$$

This shows that the torque will pulsate at angular frequencies of 4ω and higher. At the same time it can be seen that only the odd harmonics in the spatial distribution of B_r will contribute to the pulsation of the torque.

If the winding is distributed, the calculation of the torque is more complicated. *Fig. 5* shows the variation around the circum-
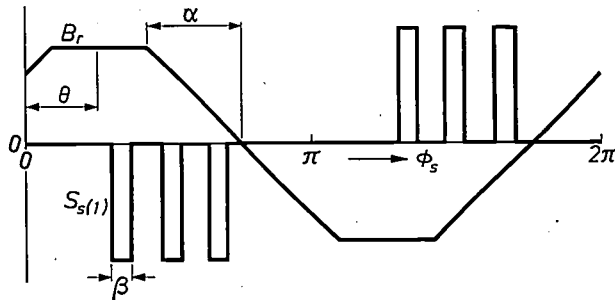
Fig. 5. The variation of the current-density distribution $S_{s(1)}$ of stator winding *1* around the stator circumference of the motor shown in fig. 4. The other curve shows a possible curve for the rotor flux density B_r around the stator circumference. The angle α characterizes this trapezoidal curve and the angle β represents the width of the slots.

ference of the stator of the current-density distribution $S_{s(1)}$, applying to stator coil *1*, for the motor of fig. 4. The 'current-density pulses' have the width of the slot opening and are reckoned as positive when the current flows 'into the paper' in fig. 4. The other curve shown in fig. 5 is an example of a possible non-sinusoidal flux-density distribution. A general method for the calculation of the torque on stator coil *1* starts with the expression, assuming certain symmetries, of both the rotor flux density and the current-density distribution in terms of a series in ϕ_s. This has already been done above for the rotor flux density, and for the current-density distribution we obtain:

$$S_{s(1)} = i_{s(1)}z_s(\phi_s) = -i_{s(1)}(z_{s1}\sin\phi_s + z_{s3}\sin 3\phi_s + \\ + z_{s5}\sin 5\phi_s + \ldots).$$

The total torque, for which the general expression

$$T_e = al\int_0^{2\pi} (S_{s(1)} + S_{s(2)})B_r a \, d\phi_s$$

is applicable, is then given by

$$T_e = -\pi a^2 l i_s \{z_{s1}B_{r1}\sin\theta_0 + z_{s3}B_{r3}\sin(3\theta_0 + 4\omega t) + \\ + z_{s5}B_{r5}\sin(5\theta_0 + 4\omega t) + \ldots\}.$$

For a purely sinusoidally distributed winding z_{s3}, z_{s5} and higher terms are zero. This cannot be achieved for all the terms. However, some degree of compensation for this can be obtained by choosing a distribution for the rotor flux that will make the remaining troublesome terms in the series for B_r very small. With the configuration shown in fig. 5, for $\alpha = \pi/3$ and $\beta = \pi/60$ radians:

$$\left|\frac{z_{s3}}{z_{s1}}\right| = 0.37 \quad \left|\frac{z_{s5}}{z_{s1}}\right| = 0.27 \quad \left|\frac{z_{s7}}{z_{s1}}\right| = 0.27$$

$$\left|\frac{B_{r3}}{B_{r1}}\right| = 0 \quad \left|\frac{B_{r5}}{B_{r1}}\right| = 0.04 \quad \left|\frac{B_{r7}}{B_{r1}}\right| = 0.02$$

The amplitudes of the terms in $3\theta_0$, $5\theta_0$ and $7\theta_0$ are thus respectively 0%, 1% and 0.5% of the maximum value of the first term.

Three-phase and single-phase machines

The machine shown in fig. 4 has two windings that together set up a current-density wave whose fundamental spatial harmonic is given by

$$S_s = i_s z_{s1}\sin(\omega t - \phi_s).$$

Here z_{s1} is the amplitude of the fundamental of the

copper-distribution function $z_s(\phi_s)$. Such a machine is called a two-phase machine. A current-density wave can also be excited by using three stator coils, separated by angles of $2\pi/3$ radians (see *fig. 6*), and supplied by a.c. currents similarly separated in phase by $2\pi/3$ radians: $i_{s(1)} = i_s\cos\omega t$, $i_{s(2)} = i_s\cos(\omega t - 2\pi/3)$, $i_{s(3)} = i_s\cos(\omega t - 4\pi/3)$. This is called a three-phase system. The fundamental of the current-density wave produced in such a three-phase machine is given by

$$S_s = \tfrac{3}{2}i_s z_{s1}\sin(\omega t - \phi_s).$$

The current-density wave around the stator circumference produces a magnetic field in the air gap. For a
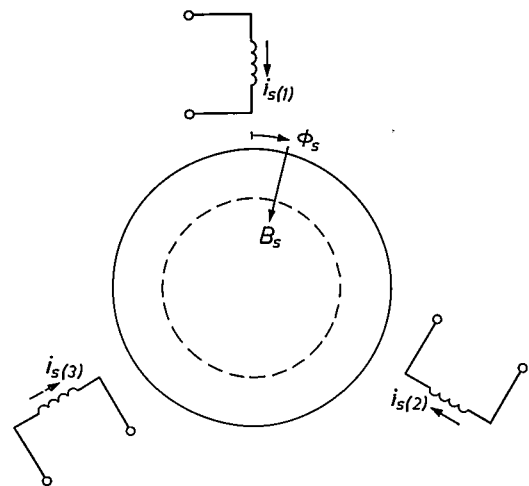
Fig. 6. Winding of a synchronous motor operating from a three-phase supply. The rotating stator field B_s is excited by the three stator coils carrying a.c. currents separated in phase by $2\pi/3$ radians.

purely sinusoidal current-density wave the magnetic field wave excited around the stator circumference (the rotating field) is also sinusoidal, provided that the air gap is uniform or the rotor consists of a material whose permeability is equal or nearly equal to that of air (e.g. permanently magnetized ferroxdure). The flux-density wave B_s is then always displaced in phase by $\pi/2$ radians with respect to the current-density wave S_s.

A special case arises in a synchronous motor when only one stator winding is energized, say winding *1*. The conditions for a current-density wave of constant amplitude and angular velocity are then no longer satisfied. The fundamental of the current-density distribution then has the form $-i_{s(1)}z_{s1}\sin\phi_s$ and is zero for $\phi_s = 0$ and $\phi_s = \pi$. This 'standing' wave can however be resolved into two travelling waves rotating in opposite directions:

$$-i_{s(1)}z_{s1}\sin\phi_s\cos\omega t = \tfrac{1}{2}i_{s(1)}z_{s1}\sin(\omega t - \phi_s) + \\ + \tfrac{1}{2}i_{s(1)}z_{s1}\sin(-\omega t - \phi_s).$$

The motor — it is called a single-phase motor — can thus rotate clockwise or anticlockwise at the synchronous speed. However, because of the current-density wave in the opposite sense the torque is not constant, but pulsates at twice the mains frequency.

Multipole machines

In all the machines so far discussed the air-gap field has a single pair of poles, and they are therefore known as two-pole rotating-field machines; the angular velocity of the rotating field is equal to the angular frequency ω of the supply mains.

A stator can also be wound in such a way that the air-gap field has four poles; alternate north and south poles are then obtained around the stator with $\pi/2$ radians between poles. *Fig. 7* shows how the stator windings are arranged in the four-pole version of the motor of fig. 3. Winding 2 is now displaced by $\pi/4$ radians with respect to winding 1, which means that the angular velocity of the rotating field, and therefore the synchronous angular velocity of the motor, is equal to half the angular frequency of the mains. The rotor is magnetized in such a way that it also has four poles.

This provides a way of reducing the speed of a synchronous motor by a factor of two. The procedure can be extended to give six-pole and eight-pole motors, or in general terms $2p$-pole motors, where p is the number of pairs of poles.

Supply from a constant-voltage source

In all the calculations given above for the torque the situation has been simplified by the assumption that the magnitude of the stator currents is given. In practice this is rarely the case. The motor is almost always connected to a source of supply whose voltage at the terminals is given; the magnitude of the currents then varies with the load on the motor.

To derive an expression for the torque in terms of the terminal voltage instead of the stator currents we must calculate the currents that flow for a given voltage at the terminals. The voltage induced in the stator winding by the alternating magnetic fields is an important quantity in these calculations. Part of this voltage will be related to the magnetic field produced by the stator winding itself and will depend upon the inductance L_s of the winding; there is also a magnetic field produced by the rotor that will induce a voltage in the stator winding when the rotor is rotating. A useful representation is given by the equivalent circuit shown in *fig. 8* for a stator winding. This equivalent circuit corresponds to the equation in complex notation:

$$V_s = (R_s + j\omega L_s)I_s + E_{sr}. \qquad (4)$$

Here V_s, I_s and E_{sr} are complex numbers whose moduli are equal to the r.m.s. values of the terminal voltage, the current and the motional e.m.f. or 'speed voltage' induced by the rotor field respectively, and with arguments equal to the appropriate phase angles. R_s is the resistance of the stator winding, and as previously ω is the angular frequency of the mains supply.

Equation (4) gives the desired relation between applied voltage and current; if we introduce the parameter $\lambda = |E_{sr}/V_s|$ then I_s can be expressed in explicit terms:

$$I_s = \frac{V_s}{R_s + j\omega L_s}\left[1 - \lambda e^{j(\theta_0 + \pi/2)}\right]. \qquad (5)$$

The parameter λ has a constant value for a synchronously operating motor; for a terminal voltage of given amplitude and frequency λ depends only on the construction of the motor. When the rotor is stationary λ in equation (5) is equal to zero; the stator current then takes the value corresponding to the impedance $R_s + j\omega L_s$ of the stator winding. The angle $\theta_0 + \pi/2$ is the phase angle of the voltage induced by the rotor field with respect to the applied voltage; θ_0 is the rotor position at the instant when the voltage across the winding reaches its maximum value. As we saw earlier, θ_0 varies with the load on the motor; equation (5) shows how I_s then varies.
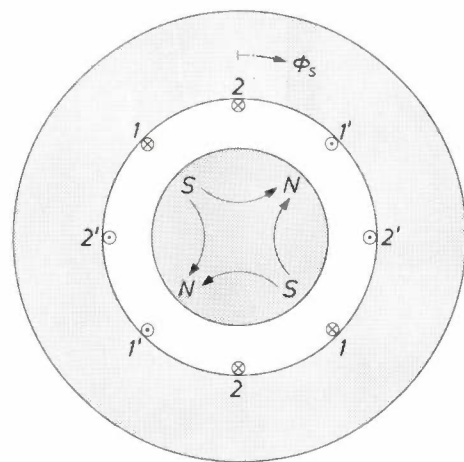
Fig. 7. Stator windings and magnetization of the rotor in a four-pole synchronous motor. The two halves of the stator winding 1 and 2 carry the same current. In each cycle of the a.c. supply the motor makes half a revolution.
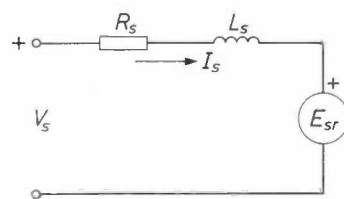
Fig. 8. Equivalent circuit of a synchronous motor with stator resistance R_s and stator inductance L_s. V_s supply voltage. I_s current in the stator winding. The e.m.f. E_{sr} induced by the field of the rotating rotor is represented by a constant-voltage a.c. source.

The equations (4) and (5) apply to each stator winding separately. The calculation of the torque, for example for the motor shown in fig. 3 with two windings, is most readily performed by setting up a power balance. The mechanical power $T_e\omega$ can then be set equal to the electrical power taken up by the two stator windings, less the ohmic losses in the windings.

The power balance can be expressed in complex notation by:

$$T_e\omega = 2\,\mathrm{Re}(I_{s(1)}{}^*V_{s(1)} - I_{s(1)}{}^*I_{s(1)}R_s) = 2\,\mathrm{Re}(I_{s(1)}{}^*E_{sr(1)}). \quad (6)$$

Here the asterisk indicates the complex conjugate. The power balance is the same for both stator windings; it is therefore sufficient to take the electrical power for a single stator winding and double it.

With the aid of the two equations (4) and (5) the current $I_{s(1)}$ and the speed voltage $E_{sr(1)}$ can be expressed in terms of the terminal voltage V_s in (6). Returning now to real quantities and writing V_s for the r.m.s. value of the terminal voltage, then from the power balance:

$$T_e = \frac{2}{\omega}\frac{\lambda V_s{}^2}{\sqrt{R_s{}^2 + \omega^2 L_s{}^2}}\{\cos(\theta_0 + \pi/2 + \chi_s) - \lambda\cos\chi_s\}\,.$$

Here $\chi_s = \arctan \omega L_s/R_s$, the phase angle of the stator impedance.

If R_s is much smaller than ωL_s the stator resistance can be neglected in the expression for the torque. A convenient expression for the torque at a given r.m.s. value V_s of the terminal voltage can then be obtained from the power balance:

$$T_e = -\frac{2\lambda V_s{}^2}{\omega^2 L_s}\cos\theta_0\,. \quad (7)$$

For motor operation T_e is positive and therefore

$$-\frac{3\pi}{2} < \theta_0 < -\frac{\pi}{2};$$

for stable operation however θ_0 must not be smaller than $-\pi$, since as soon as the motor falls behind by more than π radians the torque decreases and the motor falls out of step. Note that the angle θ_0 is the 'phase angle' of the rotor with respect to the a.c. voltage across winding 1, whereas in (3) θ_0 is a phase angle with respect to the current in winding 1.

Synchronous motors with non-uniform air gap

There are also synchronous motors that do not have a cylindrical stator bore but have salient poles (*fig. 9a*). In these motors the rotor can be subject to a torque even when there is no current in the stator winding. This is because it tries to take up a position in which the reluctance (the 'magnetic resistance') to the flux of the permanent magnet is at a minimum. The torque involved here is a stator-reluctance torque. In a motor like the one in fig. 9a this makes no contribution to the

energy conversion, since on rotation its average value is equal to zero.

If the rotor is non-cylindrical (fig. 9b) instead of the stator and also has a magnetic permeability greater than that of air, then a rotor-reluctance torque can be produced under the influence of the stator field. The operation of the reluctance motor is based upon this torque.

In the treatment of both types of motor it is probably best to start from energy considerations, since we can then obtain some insight into their characteristics without previous knowledge of the complicated variation of the magnetic field. Also, a direct calculation of the electromagnetic forces acting on the stator in the case of fig. 9a would not only have to include the Lorentz force on the conductors of the stator winding: it would also have to include the forces operating on the stator core. Such a calculation would be a complicated exercise. To make the energy considerations as general as possible it will be assumed that we have a kind of motor that combines the characteristic features of both types; with stator and rotor both non-cylindrical, and with a permanent-magnet rotor that nevertheless has a higher permeability than that of air (fig. 9c). In this motor we are concerned with energy in various forms: the applied electrical energy, which in a short time Δt is equal to $v_s i_s\Delta t$, the energy dissipated in the resistance of the stator winding, equal to $R_s i_s{}^2\Delta t$, the magnetic-field energy present in the motor W_m, and the mechanical energy $T_e\Delta\theta$ supplied by the shaft to the load during a rotation $\Delta\theta$. The applied energy is equal to the sum of the dissipated and supplied energy and the change in the magnetic-field energy; in differential form:

$$v_s i_s\mathrm{d}t = R_s i_s{}^2\mathrm{d}t + \mathrm{d}W_m + T_e\mathrm{d}\theta.$$

Here W_m is a function of the stator current i_s and the angular position θ of the rotor.

To derive the torque from this energy balance we make use of an expression for the voltage v_s across the stator winding:

$$v_s = R_s i_s + \frac{\mathrm{d}\Phi_s}{\mathrm{d}t}.$$

where Φ_s is the total magnetic flux linked with the stator winding. To a first approximation this linked flux can be equated to the product of the total magnetic flux through a stator pole and the total number of turns. Φ_s is also a function of both i_s and θ.

Substituting in the energy balance it follows that

$$T_e = i_s\frac{\partial\Phi_s}{\partial\theta} - \frac{\partial W_m}{\partial\theta}\,. \quad (8)$$

Both terms on the right-hand side are connected with

Fig. 9. *a*) Synchronous motor with a non-uniform air gap and salient poles. The permanent-magnet rotor has a preferential position, even when the stator winding is not energized. *b*) Non-cylindrical rotor in a cylindrical stator. If the permeability of the rotor is greater than that of air, it takes up a preferential position parallel to the existing stator field. *c*) A scheme for a motor combining the characteristics of (*a*) and (*b*).
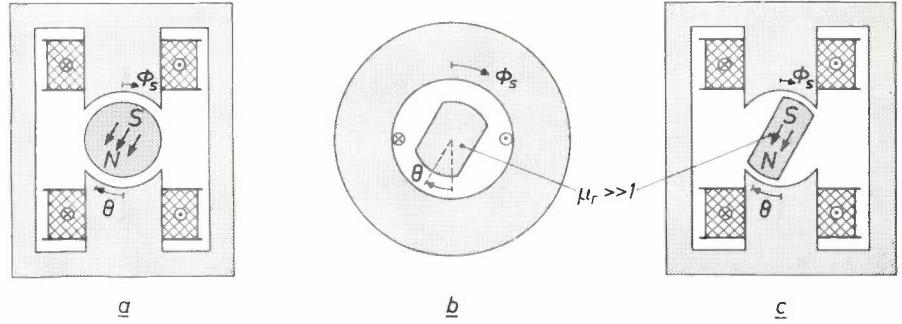
the magnetic field. To be able to draw the correct conclusions for the various types of motor from this equation, we must separate the flux Φ_s linked with the stator winding and also the magnetic-field energy W_m into a contribution from the stator current, characterized by the inductance L_s, and a contribution from the rotor. Both components vary with the angular position θ of the rotor. We then have:

$$\Phi_s = L_s(\theta)i_s + \Phi_{sr}(\theta),$$
$$W_m = \tfrac{1}{2} L_s(\theta)i_s^2 + W_{mr}(\theta).$$

Substituting this in (8) gives the general expression for the torque

$$T_e = i_s \frac{d\Phi_{sr}(\theta)}{d\theta} - \frac{dW_{mr}(\theta)}{d\theta} + \tfrac{1}{2} i_s^2 \frac{dL_s(\theta)}{d\theta}. \quad (9)$$

The first term is equal to the calculated torque in equation (6) for the motor with cylindrical stator and rotor. However, in the general case there are extra terms; the second term is related to the stator reluctance, the third to the rotor reluctance.

Single-phase synchronous motor with non-cylindrical stator

We now confine ourselves to the case shown in fig. 9*a*, of a motor with a cylindrical rotor magnetized along a diameter and a non-cylindrical stator. This is a single-phase synchronous motor, and therefore gives the pulsating torque peculiar to this type. This is no great handicap in applications such as small domestic equipment, for which such motors are ideally suited because of their simple construction. In the design however full account must be taken of the notable instabilities that can appear in this motor [4].

Since the motor has a cylindrical rotor the inductance L_s of the stator winding is not dependent on the angular position θ of the rotor, and the last term in expression (9) for the torque therefore disappears. It can be shown that the quantities appearing in the other two terms depend on θ as follows:

$$\Phi_{sr}(\theta) = \hat{\Phi}_{sr} \cos(\theta + \alpha),$$
$$W_{mr}(\theta) = W_1 - W_2 \cos 2(\theta + \beta), \quad W_1 > W_2.$$

The angles α and β are determined by the geometry of

the stator poles; they are both zero if the stator is symmetrical about the plane $\phi_s = 0$, as in fig. 9*a*. In general:

$$T_e = -i_s\hat{\Phi}_{sr} \sin(\theta + \alpha) - 2W_2 \sin 2(\theta + \beta). \quad (10)$$

The second term represents the torque for zero stator current; this reluctance or 'detent' torque can be felt on turning the shaft by hand. It pulls the rotor to one of the two stable rest positions ($\theta = -\beta$ or $\theta = -\beta + \pi$) where it remains. If the current is then switched on, a torque is only produced if α is not equal to β. In the configuration shown in fig. 9*a* ($\alpha = \beta = 0$) this torque does not arise and the motor will not run. By giving the pole arcs of the pole pieces an asymmetrical shape α can be made unequal to β, thus giving a self-starting motor.

The torque pulsations found at twice the mains frequency in this motor originate from two sources. The first is the backward-running rotating field already mentioned in the single-phase motor with cylindrical stator; the second is the stator reluctance. This appears in the expression that we find for the torque when we express the time dependence of $i_s = \hat{i}_s \cos \omega t$ and $\theta = \omega t + \theta_0$ in (10):

$$T_e = -\tfrac{1}{2} \hat{i}_s \hat{\Phi}_{sr} \{\sin(2\omega t + \theta_0 + \alpha) + \sin(\theta_0 + \alpha)\} + $$
$$-2W_2 \sin 2(\omega t + \theta_0 + \beta).$$

Reluctance motor

Reluctance motors are characterized by a non-cylindrical rotor of high-permeability material. The stator bore can be either cylindrical or non-cylindrical. *Fig. 10* shows a single-phase two-pole reluctance motor.

The rotor is not a permanent magnet and the quantities Φ_{sr} and W_{mr} in expression (9) for the torque now disappear. The simple expression remaining is

$$T_e = \tfrac{1}{2} i_s^2 \frac{dL_s(\theta)}{d\theta}.$$

The variation of the stator inductance L_s with the rotor position θ is not in general easy to calculate. As a first

[4] These have been investigated at the Philips Aachen laboratories; see: H. Schemmann, Stability of small single-phase synchronous motors; this issue, p. 235.

approximation for the case shown here let us assume that

$$L_s = L_1 + L_2 \cos 2\theta, \quad L_1 > L_2.$$

From this we have for the torque:

$$T_e = -i_s^2 L_2 \sin 2\theta.$$

This torque is zero for four positions of the rotor; the two positions perpendicular to the stator field ($\theta = \pi/2$ and $\theta = 3\pi/2$) are unstable and the two parallel to the stator field ($\theta = 0$ and $\theta = \pi$) are stable. When the stator winding is energized with d.c. or a.c. current the rotor takes up one of the last two positions; the motor thus has a positioning effect but is not self-starting, even with a.c. current.

Motor operation will only occur if the motor runs synchronously with the stator current $i_s = \hat{i}_s \cos \omega t$; the torque is then given by

$$T_e = -\tfrac{1}{4} i_s^2 L_2 \{ 2 \sin (2\omega t + 2\theta_0) + \\ + \sin (4\omega t + 2\theta_0) + \sin 2\theta_0 \}.$$

This is a pulsating torque with the average value:

$$\langle T_e \rangle = -\tfrac{1}{4} \hat{i}_s^2 L_2 \sin 2\theta. \tag{11}$$

The positioning effect of the reluctance motor is made use of in a stepping motor of the reluctance type. An example with four stator poles and five rotor lobes is shown in *fig. 11*. The four stator windings are energized in succession with d.c. current; each time the next stator coil is energized the motor performs a step of $(\tfrac{1}{4} - \tfrac{1}{5}) \times 360° = 18°$. The control pulses are produced electronically. *Fig. 12* is a photograph of a stepping motor with eight lobes on the rotor and twelve poles on the stator. In this motor four poles spaced by 90° are energized at a time; the stepping angle is 15°.

Hysteresis motor

The hysteresis motor is a special case of the synchronous motor, with the notable feature that its average torque can differ from zero at all speeds, and not just at the synchronous speed. This has the practical advantage that no special provisions have to be made for starting. If the load torque remains less than the maximum motor torque the motor runs synchronously.

We shall now look more closely at the hysteresis motor, starting with *fig. 13*, which shows a two-phase hysteresis motor. The two stator windings are distributed sinusoidally around the circumferences of the stator bore; the rotor consists of a homogeneous cylinder of a material such as cobalt steel. Let us first assume that one of the windings is energized with a d.c. current. A stationary stator field is then produced, which

magnetizes the rotor homogeneously. If the rotor is now made to rotate by external means at an angular velocity ω_r, every element of the rotor is exposed to a rotating magnetic field. The field thus rotates $\omega_r/2\pi$ times per second.
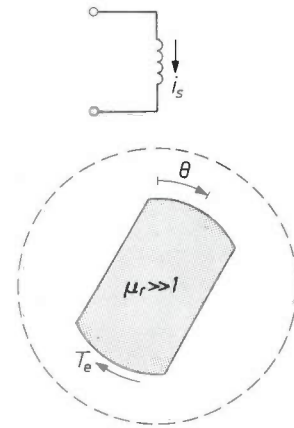
Fig. 10. Reluctance motor. The non-cylindrical rotor has a permeability much greater than that of air.
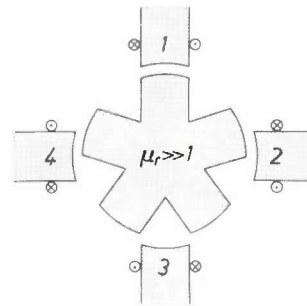
Fig. 11. Principle of the reluctance stepping motor. The four stator coils *1* to *4* are successively energized with d.c. current pulses; after a cycle of four pulses the rotor has rotated one lobe, and hence by a fifth of a revolution.
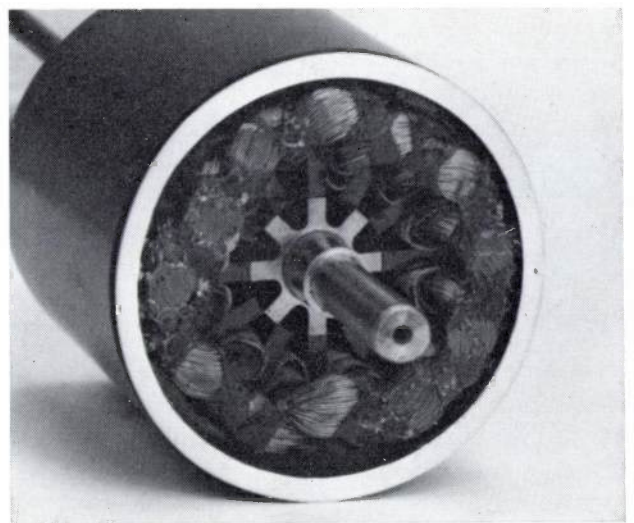
Fig. 12. Reluctance stepping motor. There are eight lobes on the rotor and twelve poles on the stator. Four stator poles spaced by 90° are energized at a time; the stepping angle is 15°.

In most hysteresis motors the rotor consists of a cylinder with marked hysteresis properties, mounted around a core that can be ferromagnetic or not. Here it is more difficult to describe the hysteresis, since it can no longer be assumed that each rotor element is subjected to a rotating field of constant amplitude and angular velocity. We shall not go further into this here.

The existence of hysteresis implies that the magnetization of each element is delayed in direction with respect to the inducing stator field. The angular difference is independent of the speed of rotation, and is determined only by the magnitude of the stator field and the hysteresis properties of the rotor material. Because of this angular difference a braking torque T_e is experienced when the rotor is driven; this braking torque is also independent of the speed of rotation. *Fig. 14a* shows the torque as a function of the speed of rotation: the torque-speed characteristic. Mechanical power $-T_e\omega_r$ is thus applied to the motor; this is completely converted into the hysteresis losses P_{hyst} associated with the rotational magnetization of the rotor material:

$$-T_e\omega_r = P_{hyst}.$$

The hysteresis losses are proportional to the speed, since in each rotor element the direction of magnetization rotates once per revolution of the rotor; it follows from this that the braking torque is independent of the speed.

In motor operation the stator is energized in such a way as to produce a stator field that rotates once for each period of the supply current. If the angular speed ω_r of the rotor is smaller than that of this rotating field, then the rotor will rotate in the opposite direction to the field and a similar situation arises as described above for the stationary stator field. If ω is the angular velocity of the rotating field, the direction of magnetization rotates in the rotor material $(\omega - \omega_r)/2\pi$ times per second and we have:

$$T_e(\omega - \omega_r) = P_{hyst}. \qquad (12)$$

The torque-speed characteristic is now that of fig. 14b.

When the motor is running synchronously the torque can take any value between the extreme values shown in fig. 14b. The hysteresis motor then behaves rather like a synchronous motor with a permanent-magnet rotor. This means that the angle at which the rotor magnetization follows the rotating stator field now depends upon the load. As soon as this angle tends to become larger than the angle resulting from the hysteresis, the rotor starts to 'slip' and falls below the synchronous speed. The angle between the magnetization and the rotating field then remains at the fixed value determined by the hysteresis.

In practice there are deviations from the rectangular curve shown in fig. 14b. These are due to the slots in the stator bore for the winding, the associated higher harmonics in the current-distribution pattern, and the eddy currents in the rotor body.

A hysteresis motor wound for two-phase operation as in fig. 13 can easily be modified for operation from the ordinary single-phase mains. Usually one of the windings (the main winding) is connected directly to the mains, while the other (the auxiliary winding) is connected through a capacitor. The effect of the capacitor is to advance the phase of the current in the auxiliary winding with respect to the current in the main winding. This phase difference depends to some extent on the speed of the motor. By carefully choosing the element values in the auxiliary circuit the phase differences can be made equal to $\pi/2$ at for example zero or nominal speed. The motor will then operate like a true two-phase motor at this operating point. In the first case we have a 'starting capacitor', in the second a 'running capacitor'.
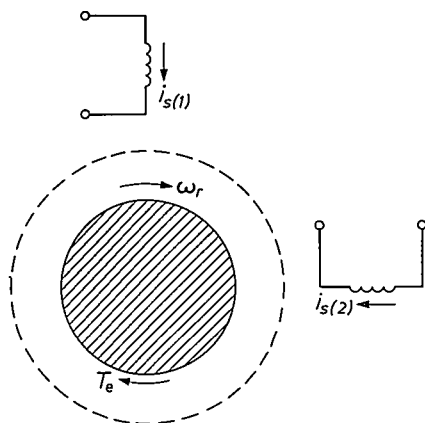
Fig. 13. Two-phase hysteresis motor. The rotor is a homogeneous cylinder of material with a high magnetic hysteresis, such as cobalt steel. Because of the hysteresis the direction of the magnetization in the rotor always lags behind that of the stator field, so that a torque is exercised on the rotor.
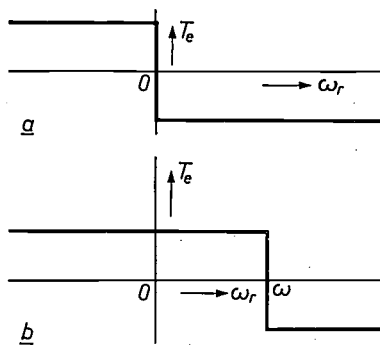
Fig. 14. Torque-speed characteristic of a hysteresis motor when the stator windings are energized with (a) d.c. currents, (b) a.c. currents.

II. Asynchronous motors (induction motors)

So far we have been concerned with motors whose operation does not depend upon currents in the rotor. Now we shall look at motors that do depend on such currents, and we shall begin with the motors in which these rotor currents are not supplied by an external source but are induced by the stator field. These are the induction motors. An induction motor has a rotor with closed current circuits. If these are not accessible from outside it is a squirrel-cage rotor machine. Sometimes the current circuits are given external access via slip rings for control purposes; then we have a slip-ring-rotor (or 'wound-rotor') machine.

When an induction motor runs synchronously with the rotating stator field, the rotor is stationary with respect to the field and no rotor currents are induced; the torque is then zero. In normal operation the motor 'slips' with respect to the rotating field; it is because of this slip that currents are induced in the rotor and a torque is produced. For this reason these motors are also known as asynchronous motors. Within certain limits they can suit their speed to the torque required, which in many applications is an advantage over the rather inflexible behaviour of the synchronous motor. There is also a torque when the rotor is stationary, so that induction motors are self-starting. These characteristics, and the little maintenance required — since current does not have to be supplied to the rotor via sliding contacts subject to wear — make the induction motor a tough and reliable drive unit that gives useful service in many applications.

To give a simple treatment, the torque will first be derived for known stator currents, and then for a known terminal voltage. The torque is a function of the speed; in practice the torque-speed characteristic is very useful in assessing the performance of a motor.

Derivation of the torque

With stator currents known

We shall assume a two-phase, two-pole induction motor whose rotor carries two short-circuited diametral coils (*fig. 15*). The stator windings are distributed around the stator circumference in such a way that higher harmonics in the stator field can be neglected. No account will be taken here of iron losses due to eddy currents and hysteresis. There are two ways of calculating the torque from the Lorentz forces. One is based on the forces experienced by the stator conductors as a result of the magnetic field excited by the rotor currents, and the other is based on the forces experi-

enced by the rotor conductors as a result of the stator field. We shall use the second method.

The stator field is similar to the rotating stator field in the synchronous motor and is given by

$$B_s(\phi_s) = \hat{B}_s \cos(\omega t - \phi_s).$$

The new feature is the calculation of the rotor currents. These have a frequency proportional to the difference between the angular velocities of the stator field and the rotor, and hence proportional to the slip. We define the slip s as

$$s = \frac{\omega - \omega_r}{\omega}.$$

The rotor currents $i_{r(1)}$ and $i_{r(2)}$ can be calculated with the aid of the voltage equation for the closed rotor circuit. For rotor coil I the equation becomes:

$$0 = R_r i_{r(1)} + L_r \frac{di_{r(1)}}{dt} + \frac{d\Phi_{rs(1)}}{dt}. \tag{13}$$

Here R_r is the resistance of the coil, L_r is its inductance and $\Phi_{rs(1)}$ is the part of the stator flux linked with rotor coil I. The magnitude of Φ_{rs} is proportional to the mutual inductance M.
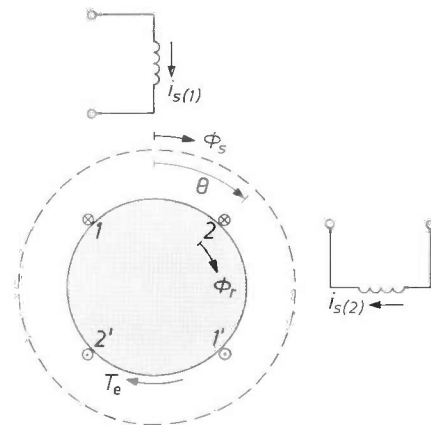
Fig. 15. Two-phase, two-pole induction motor. The stator has two sinusoidally distributed windings, which together excite a rotating field. The rotor has two short-circuited diametral coils $1,1'$ and $2,2'$ in which currents are induced.

When the rotor currents are known, the torque can be found as the sum of the torques on the two coils:

$$T_e = 2bN_r l\{i_{r(1)}B_s(\theta + \pi/2) + i_{r(2)}B_s(\theta + \pi)\},$$

where b is the radius of the rotor, l is its length and N_r is the number of turns on the rotor coil.

The stator flux linked with the rotor coil *1* is

$$\Phi_{\mathrm{rs}(1)} = 2blN_{\mathrm{r}}\hat{B}_{\mathrm{s}}\cos{(\omega t - \theta)}.$$

We can also write:

$$\Phi_{\mathrm{rs}(1)} = \hat{\imath}_{\mathrm{s}}\hat{M}\cos{(\omega t - \theta)},$$

where

$$\hat{M} = 2blN_{\mathrm{r}}\frac{\hat{B}_{\mathrm{s}}}{\hat{\imath}_{\mathrm{s}}}$$

is a quantity entirely determined by the construction of the motor. If we put

$$\theta = \omega_{\mathrm{r}}t + \theta_0 = (1-s)\omega t + \theta_0,$$

then we have

$$\Phi_{\mathrm{rs}(1)} = \hat{\imath}_{\mathrm{s}}\hat{M}\cos{(s\omega t - \theta_0)}.$$

When this is substituted in the differential equation (13), we obtain a solution

$$\hat{\imath}_{\mathrm{r}(1)} = \frac{s\omega\hat{\imath}_{\mathrm{s}}\hat{M}\sin{(s\omega t - \theta_0 - \chi_{\mathrm{r}})}}{\sqrt{R_{\mathrm{r}}^2 + (s\omega L_{\mathrm{r}})^2}}, \qquad (14)$$

where $\chi_{\mathrm{r}} = \arctan{(s\omega L_{\mathrm{r}}/R_{\mathrm{r}})}$ is the phase angle of the impedance of the rotor coil at the angular frequency $s\omega$ of the rotor currents. Since the current in rotor coil *2* lags behind the current in coil *1* by $\pi/2$ radians,

$$i_{\mathrm{r}(2)} = \frac{-s\omega\hat{\imath}_{\mathrm{s}}\hat{M}\cos{(s\omega t - \theta_0 - \chi_{\mathrm{r}})}}{\sqrt{R_{\mathrm{r}}^2 + (s\omega L_{\mathrm{r}})^2}}. \qquad (15)$$

The following expression can be shown to apply for the torque of the induction motor for known stator currents i_{s}:

$$T_{\mathrm{e}} = \frac{\hat{\imath}_{\mathrm{s}}^2\hat{M}^2}{L_{\mathrm{r}}}\ \frac{1}{R_{\mathrm{r}}/s\omega L_{\mathrm{r}} + s\omega L_{\mathrm{r}}/R_{\mathrm{r}}}. \qquad (16)$$

The first factor of the right-hand side contains only motor constants besides the stator current; the second factor varies with the slip and has a maximum value at $s\omega L_{\mathrm{r}} = R_{\mathrm{r}}$. This maximum torque

$$T_{\max} = \hat{\imath}_{\mathrm{s}}^2\hat{M}^2/2L_{\mathrm{r}},$$

is known as the pull-out torque and the corresponding slip $s_{\max} = R_{\mathrm{r}}/\omega L_{\mathrm{r}}$ as the pull-out slip. If the load torque rises to a value higher than the pull-out torque the speed suddenly falls and the motor comes to rest — this is the reason for the name 'pull-out torque'. On altering the resistance R_{r} of the rotor circuit, the pull-out slip changes, but not the pull-out torque. This effect can be applied in an induction motor that will give the maximum torque at a desired speed. In this kind of motor the rotor windings are brought out to slip rings, and an external control resistor completes the circuit, via brushes. This is a slip-ring-rotor or wound-rotor machine. Varying the control resistor gives different shapes of torque-speed characteristics; see *fig. 16*. The starting behaviour can also be varied in this way.

With terminal voltage known

In the derivation above of the torque of the synchronous motor supplied from a constant-voltage source, the quickest method was to set up a power balance: the electrical power supplied was set equal to the mechanical power supplied plus the heat produced in the windings. This is again the quickest method here. It will also be convenient to make use of the similarity between the induction motor and a transformer. The stator windings can be compared with the primary winding of a transformer connected to the mains, and the rotor with the secondary winding. If the rotor is stationary the induction motor is effectively a transformer with a short-circuited secondary. Once the motor starts to rotate this situation changes; a load resistance appears in the secondary circuit to represent the mechanical load (*fig. 17*). This resistance varies with the slip and is equal to $R_{\mathrm{r}}(1-s)/s$.

This expression can be derived by starting from the voltage equation for the stator windings. For winding *1*:

$$v_{\mathrm{s}(1)} = R_{\mathrm{s}}i_{\mathrm{s}(1)} + L_{\mathrm{s}}\frac{\mathrm{d}i_{\mathrm{s}(1)}}{\mathrm{d}t} + \frac{\mathrm{d}\Phi_{\mathrm{s}(1)\mathrm{r}}}{\mathrm{d}t}.$$

Here $\Phi_{\mathrm{s}(1)\mathrm{r}}$ is the magnetic flux originating from the rotor cur-
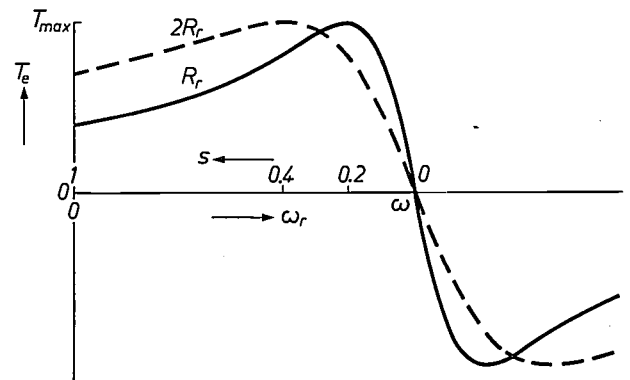
Fig. 16. Torque-speed characteristic of an induction motor. The horizontal axis indicates the slip *s* as well as the angular velocity ω_{r} of the rotor. The angular velocity of the rotating stator field is ω, and the maximum torque (the pull-out torque) of the motor is $T_{\max}$. The slip at which the pull-out torque is delivered can be made larger (dashed curve) by increasing the resistance of the rotor circuit. To enable the characteristic to be varied in this way slip rings are sometimes provided to allow a control resistor to be included in the rotor circuit. If the resistance of the rotor circuit is increased in this way from R_{r} to $2R_{\mathrm{r}}$, the pull-out slip increases from 0.2 to 0.4 in the diagram.
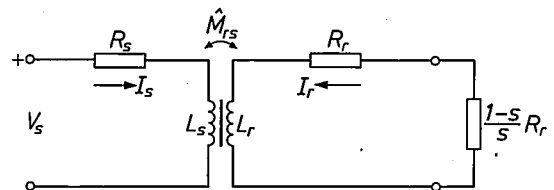
Fig. 17. Equivalent circuit for the induction motor. This is similar to the equivalent circuit of a transformer of mutual inductance $\hat{M}$; the primary side, connected to the mains supply voltage V_{s}, also includes the inductance L_{s} and the resistance R_{s} of the stator windings. The secondary circuit includes the corresponding quantities for the rotor circuit and a load resistance $R_{\mathrm{r}}(1-s)/s$, representing the mechanical load.

rents and linked with stator coil *l*. The magnetic field of the rotor is a rotating field; its distribution around the circumference of the rotor depends on the construction of the motor and is usually not sinusoidal. However, for an approximately sinusoidal current-density wave around the stator the only component of this distribution of any importance in calculating the torque is the fundamental. This is given by:

$$B_r = \hat{B}_r \sin (s\omega t - \theta_0 - \chi_r - \phi_r),$$

as can be shown from equations (14) and (15). The field therefore rotates around the rotor at an angular frequency $s\omega$. Since the rotor is itself rotating at an angular velocity $(1-s)\omega$, the angular velocity of the rotor field with respect to the stator is equal to ω:

$$B_r = \hat{B}_r \sin (\omega t - \chi_r - \phi_s).$$

This is also the angular velocity of the stator current-density wave, and in fact the two angular velocities must be equal if the average value of the torque is to differ from zero.

This equation corresponds to the equation for a transformer with a resistance R_r/s in the secondary circuit. Since the resistance of the rotor circuit is R_r there is a load resistance equal to

$$\frac{R_r}{s} - R_r = \frac{1-s}{s} R_r .$$

The power taken up in this resistance, multiplied by 2 since there are two rotor coils, is equal to the mechanical power supplied:

$$T_e(1-s)\omega = 2 \frac{1-s}{s} R_r I_r^2 .$$

(I_r is the r.m.s. value of the rotor current.) From this it follows that the torque is equal to

$$T_e = \frac{2R_r I_r^2}{s\omega} .$$
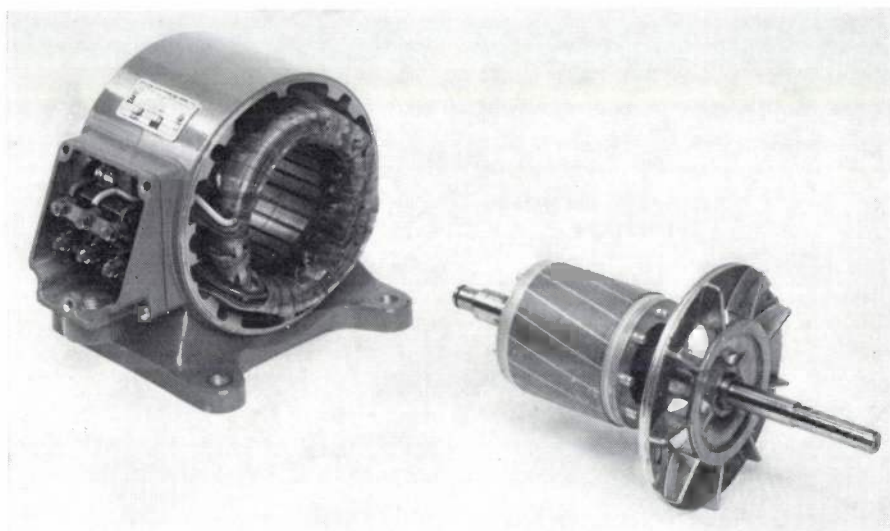
Fig. 18. Stator and rotor of a 50 watt induction motor. The rotor is a squirrel-cage rotor; bars and end rings are cast as a single piece of aluminium (together with the fan wheel, which provides internal air cooling).

The flux $\Phi_{s(1)r}$ is proportional to the amplitude of B_r at the stator coil *l*, where $\phi_s = 0$. Making use again of the mutual inductance M we have:

$$\Phi_{s(1)r} = \hat{i}_r \hat{M} \sin (\omega t - \chi_r),$$

and for the voltage equation:

$$v_{s(1)} = R_s i_{s(1)} + L_s \frac{d i_{s(1)}}{dt} + \hat{M} \frac{d}{dt} \hat{i}_r \sin (\omega t - \chi_r) .$$

From (14) and (15) we see that

$$\hat{i}_r = \frac{\omega \hat{i}_s \hat{M}}{\sqrt{(R_r/s)^2 + (\omega L_r)^2}} ,$$

which can be substituted in the voltage equation to eliminate $\hat{i}_r$. The equation then becomes, in complex notation:

$$V_{s(1)} = (R_s + j\omega L_s - \frac{\omega^2 \hat{M}^2}{R_r/s + j\omega L_r}) I_{s(1)},$$

where as before the modulus of $V_{s(1)}$ and $I_{s(1)}$ indicates the r.m.s. value.

The rotor current I_r can be expressed in terms of the stator voltage. This gives an expression for the torque that is fairly complicated; it can be simplified if we assume that the stator resistance is negligible compared with the stator reactance, as we did for the synchronous motor. We also introduce the leakage coefficient σ, frequently used in transformer analysis:

$$\sigma = 1 - \frac{\hat{M}^2}{L_s L_r} .$$

This is a measure of the coupling between primary and secondary: it approaches zero for tight coupling and unity for weak coupling. In practice σ has a value between 0 and 0.2 for induction motors and is a constant of the motor. We now obtain for the pull-out torque and slip:

$$T_{max} = \frac{1-\sigma}{\sigma} \frac{V_s^2}{\omega^2 L_s},$$

$$s_{max} = \frac{R_r}{\sigma \omega L_r},$$

and for the torque in general:

$$T_e = \frac{2T_{max}}{s/s_{max} + s_{max}/s}.$$

Three kinds of rotor

Several kinds of rotor can be used for an induction motor. The slip-ring rotor (or wound rotor) was discussed earlier; other important types are the squirrel-cage rotor and the solid rotor.

In the squirrel-cage rotor conducting bars embedded in slots in the rotor iron are short-circuited together at their ends by two conducting rings (*fig. 18*). This sturdy and relatively simple type of motor is widely used.

The solid rotor consists of a homogeneous steel cylinder or ring. Here the rotor material acts at the same time as an electrical and a magnetic conductor. It is obviously the simplest and sturdiest kind of rotor imaginable. This kind of rotor is sometimes used in high-speed induction motors operated from a high-frequency supply [*].

If the skin effect is present, as in the solid rotor and in some types of squirrel-cage rotor, the rotor resistance R_r and the rotor inductance L_r become functions of the slip. The torque-speed characteristic of these motors then differs from the curves shown in fig. 16.

Single-phase induction motor

An induction motor will also run with just one stator winding; however, such a motor has no starting torque. This can be seen as follows. The current-density distribution of the stator winding can be resolved into two waves, one rotating clockwise, and the other rotating anticlockwise. Each wave produces a torque on the rotor. *Fig. 19* shows the corresponding torque-speed characteristics. The resultant torque-speed characteristic passes through the origin. Once the motor is rotating motor operation is possible, in either direction; the torque pulsates at twice the mains frequency.

Most single-phase induction motors are provided with an auxiliary winding, which is connected in either temporarily for starting or permanently to improve normal operation. The required phase difference be-

tween the current in the main winding and the current in the auxiliary winding can be obtained by including a capacitance in series with the auxiliary winding [5] (see also p. 225), or giving the auxiliary winding a relatively high resistance.

A special type of single-phase induction motor with an auxiliary winding is the 'shaded-pole' motor. The construction of this motor is shown in the schematic diagram of *fig. 20*. The main winding consists of two coils wound on the stator poles and connected in series or parallel. The auxiliary winding merely consists of one or more short-circuited turns or shading coils SC surrounding a portion of the pole. Currents flow in these coils even when the rotor is stationary, since they are magnetically coupled to the main winding. These currents lag behind the main current and their fields
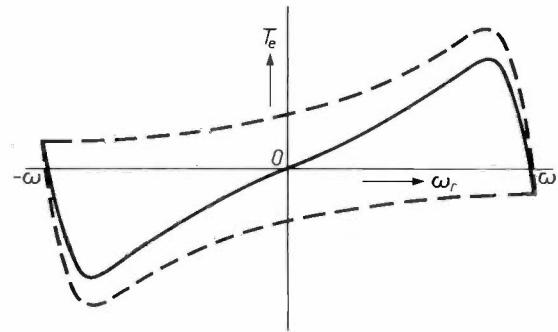
Fig. 19. Torque-speed characteristic of a single-phase induction motor. This has a pulsating stator field, which can be resolved into two rotating components, one clockwise and the other anti-clockwise. The two rotating components produce two opposing torque-speed characteristics (dashed curves); the resultant curve is the torque-speed characteristic of the motor (solid curve).
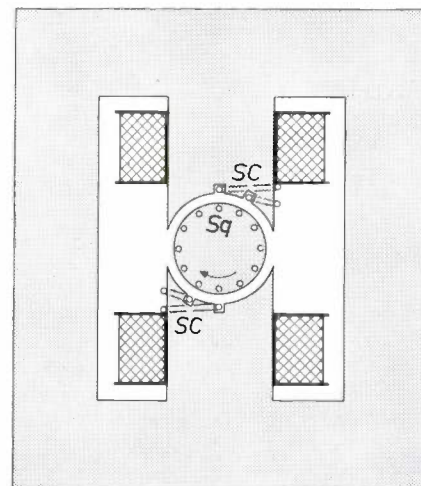
Fig. 20. Single-phase induction motor of the 'shaded-pole' type. This kind of motor is fitted with auxiliary windings SC, the 'shading coils'. The currents induced in these produce a local magnetic field that lags behind the main field; this gives rise to a primitive kind of rotating field in the air gap, so that the motor is self-starting. Sq squirrel-cage rotor.

[*] We intend to include an article on high-speed induction motors in the third issue on electric motors. (*Ed.*)

[5] See the article by K. Rennicke about speed control of capacitor motors, to appear in the second issue on electric motors.

are displaced in position with respect to the main field. The arrangement forms in effect a primitive kind of rotating-field motor, which is not only self-starting but has a better running characteristic than it would have without the shaded poles. *Fig. 21* is a photograph of a shaded-pole motor with three shading coils.
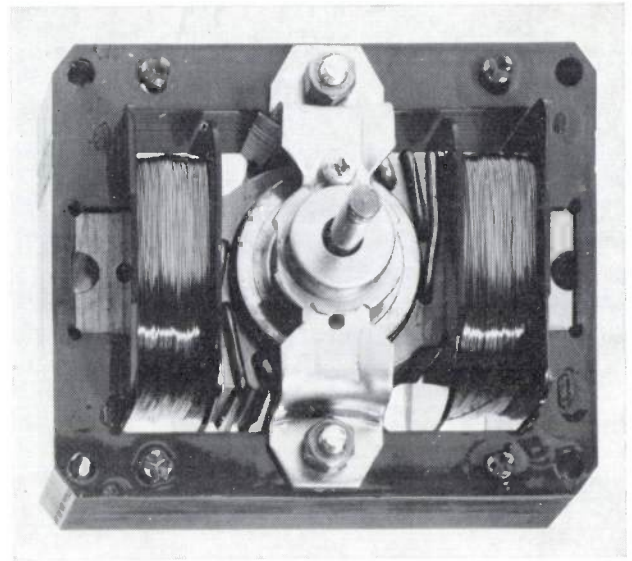
Fig. 21. Shaded-pole motor with three shading coils; this motor is mainly used in record players. The squirrel-cage rotor has copper bars and end rings.

III. Commutator motors

The principle of the commutator motor — the type that has to be used when the supply is a d.c. source — wil be explained with the aid of *fig. 22*, which shows a commutator motor in its simplest form. A rectangular loop of wire can rotate in a magnetic field B; the axis of rotation is at right angles to the direction of B. If a current I flows in the loop it will be subject to a torque; this torque varies as the sine of the angle of rotation θ and is therefore zero when averaged over a complete revolution. To obtain a torque whose average value is different from zero the current is reversed at the positions $\theta = -\pi/2$ and $\theta = \pi/2$; the torque then varies as shown in *fig. 23*. The usual way of reversing the
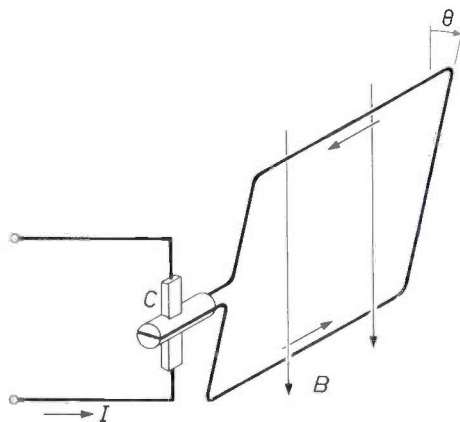
current is by means of a mechanical commutator on the shaft of the motor; the commutator in fig. 22 consists of two segments and two brushes.

The average value of the torque is of course much larger if the rotor is arranged to carry a large number of coil loops which are all supplied with current in the way just described. Moreover, the variations in the torque will be proportionally reduced. A winding of this type is the lap winding. The principle of this winding is illustrated in the developed diagram shown in *fig. 24* of a simple lap winding with four diametral coils. These coils are arranged in practice in slots in the circumference of a laminated cylindrical core. Each slot carries two coil sides, one above the other in the slot. All the coils are connected in series, while the four connecting points are each connected to a segment of the commutator. The result is that the circuit appearing between the brushes is an armature winding consisting of two paths in parallel, the 'armature paths'.

When a brush goes from one segment to the next, one of the coils is short-circuited by this brush. The commutation is said to be ideal if the current in the short-circuited coil varies linearly from $I/2$ to $-I/2$ or vice versa (straight-line commutation) during the short-circuiting.

The attainment of good commutation requires much careful attention, particularly in the larger motors, and special provisions such as brush-shift, commutating poles and perhaps compensating windings are quite usual. However, since we are mainly concerned here with small motors we shall not discuss these matters further.

Fig. 22. A commutator motor in its simplest form. A rectangular loop of wire can rotate in a magnetic field of flux density B. The commutator C reverses the current I in the loop at the positions $\theta = -\pi/2$ and $\theta = \pi/2$; this ensures that the torque exerted on the loop has the same sign during the complete revolution.

If a d.c. current is applied to the brushes, then each coil of the armature winding takes an a.c. current whose waveform is approximately a square wave. The commutator ensures that the angular frequency of this square wave always corresponds to the angular velocity of the rotor; it functions effectively as a frequency converter. In this case it converts the zero frequency of the d.c. current into a frequency corresponding to the rotor speed.
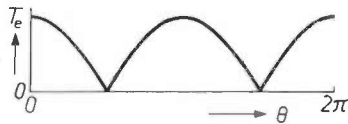
Fig. 23. The torque exerted on the loop of fig. 22, as a function of the angle θ.
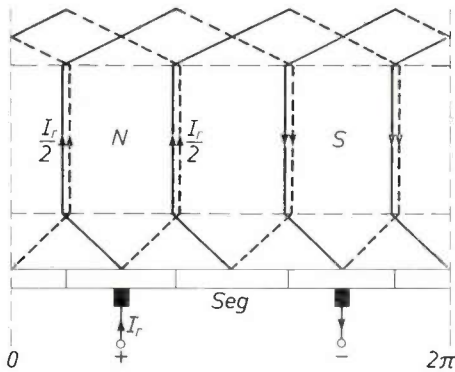
Fig. 24. Development of a lap winding consisting of four diametral coils. Each coil is uppermost in a slot on one side of the rotor (solid lines), and at the bottom of a slot on the other side (dashed lines). The connection between each pair of coils is connected to a commutator segment *Seg*. The rotor winding gives two armature paths, which appear in parallel between the brushes; half the rotor current I_r flows in each path. N and S indicate the north and south poles.

Torque and speed voltage

Fig. 25a is a schematic diagram of a commutator motor with a field winding on salient poles. In a motor of this construction the main field around the circumference of the rotor is approximately trapezoidal. If the current distribution around the rotor is uniform — which means that there must be a large number of rotor windings and hence a large number of segments — then the 'current packet' present in the stator field will always be the same and the torque will be practically constant. The general expression for this torque is

$$T_e = -b^2 l \int_0^{2\pi} S_r B_s \mathrm{d}\phi_r,$$

where as before b is the radius, l the length and S_r is the current density of the rotor. For the present it will be sufficient to derive from this the relation

$$T_e = C I_r \Phi_{rs}. \tag{17}$$

Here Φ_{rs} is the magnetic flux penetrating the rotor from a stator pole and C is a constant determined by the construction of the motor.

As soon as the rotor is in motion a voltage is induced in its conductors. The sign and magnitude of this voltage is given by the general vector relation $\boldsymbol{E} = l\boldsymbol{v} \times \boldsymbol{B}$ where $\boldsymbol{v}$ is the velocity of the conductor. In our case the magnitude is equal to $\omega_r b l B_s$, since the rotor speed at the circumference is $\omega_r b$. This voltage can be measured at the rotor terminals when there is no current in the rotor and it is externally driven. It is called the speed voltage E_r, and from the above it is proportional to $\omega_r \Phi_{rs}$:

$$E_r = C \omega_r \Phi_{rs}.$$

The proportionality constant is the same as in (17); this follows from the power balance

$$T_e \omega_r = E_r I_r,$$

which is the same as that discussed on p. 222 for the synchronous motor, since $E_r I_r$ is equal to the electrical power supplied less the copper losses

$$E_r I_r = V_r I_r - R_r I_r^2.$$

Here R_r is the rotor-circuit resistance, including the brush-contact resistance (see fig. 25b).
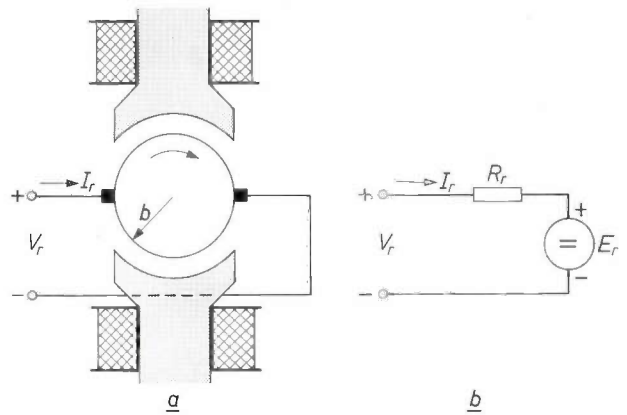
Fig. 25. *a*) D.C. commutator motor with a field winding on salient poles. V_r voltage at rotor terminals. I_r rotor current. *b*) Equivalent circuit for a d.c. motor. R_r resistance of rotor winding and commutator. E_r speed voltage.

To say more about the behaviour of the torque for varying supply current, voltage or speed, we need to know more about the motor circuit. This is because, unlike the types discussed earlier, this kind of motor has two windings that have to be connected to the source. There are three ways of doing this.

Three types of commutator motor

Commutator motors can be divided into three types with different circuits: the separately excited motor, the shunt motor and the series motor (*fig. 26*). In the separately excited motor the field winding and the rotor each have a separate supply that can be independently operated. Motors with permanent-magnet excitation can be included in this class, in which there is of course no question of controlling the main field. Permanent-magnet excitation is usually only employed in small motors.

Even though the motor with independent energization of the field winding already has appreciable scope for control, a starting resistance R_{ext} is often used in the rotor circuit, as with the other types. In larger motors the starting resistance is in fact necessary to limit the current when starting at full mains voltage; at the instant when the motor is connected to the supply no speed voltage has as yet appeared to oppose the supply voltage and with no extra resistance there would be a very large surge of current. From the equations given earlier it follows that the general relation between torque and angular velocity for the d.c. motor is

$$T_e = C\Phi_{rs} \frac{V_r - C\omega_r \Phi_{rs}}{R_r + R_{ext}}. \tag{18}$$

This relation can be altered by varying the three parameters Φ_{rs}, V_r and R_{ext}; independent variation is only possible for the separately excited motor with field winding. The effect of such independent variation of the three parameters on the torque-speed characteristic is shown diagrammatically in *fig. 27*. It is clear from this figure that the d.c. motor is basically very adaptable; it is therefore very suitable for control systems.

In the shunt and series motors the three parameters cannot be varied independently. In the shunt motor the field winding and the rotor are connected in parallel across the supply voltage V_r (fig. 26*b*). If this voltage has a constant value, Φ_{rs} is constant and the torque-

Fig. 27. The changes in the torque-speed characteristic of the commutator motor for variation of *a*) the stator flux Φ_{rs}, *b*) the rotor terminal voltage V_r and *c*) the starting resistance R_{ext} in the rotor circuit.

speed characteristic is a straight line with a negative slope that can be varied by means of R_{ext} (fig. 27*c*). In the series motor the field winding is in series with the rotor (fig. 26*c*); the main-field flux Φ_{rs} now varies with the rotor current I_r. When the loading torque is increased the speed decreases and the rotor current increases; consequently Φ_{rs} also increases at lower speeds. Curve *Se* in *fig. 28* is an example of the torque-speed characteristic of a series motor; it shows the same variation of the slope when Φ_{rs} is varied as in fig. 27*a*.

To determine the torque-speed characteristic of a shunt or series motor completely we must know the relation between the main-field flux Φ_{rs} and the excitation current I_s. This relation is only linear for small values; at higher values it is affected by magnetic saturation. The relation can be measured by driving the motor at constant speed and measuring the voltage produced at the rotor terminals as a function of the excitation current; this gives the relation since the speed voltage is proportional to the main field. The curve obtained in this way is called the

Fig. 26. The three circuit arrangements that can be used with a commutator motor. *a*) Separately excited motor; field winding and rotor are connected to two independent sources of supply. A starting resistance R_{ext} is connected in series with the rotor. Motors with permanent-magnet energization are also included under this type. *b*) Shunt motor; field winding and rotor are connected in parallel across the same supply voltage. *c*) Series motor; field winding and rotor carry the same current.

open-circuit characteristic; *fig. 29* gives a typical example. D.C. motors are usually designed to allow a small amount of magnetic saturation under nominal operating conditions.
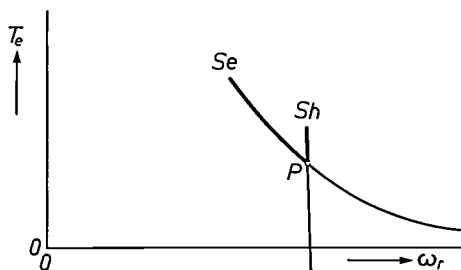
Fig. 28. Torque-speed characteristics for a shunt motor (curve *Sh*) and a series motor (curve *Se*) with the same nominal operating point *P*. When the load on the motors is greater than the nominal torque the operating point moves upwards along the curves, which in the figure are continued to the point at which the rotor current has increased to the maximum permissible value (assumed to be the same for both motors). The diagram shows that the shunt motor reaches this limit at a lower torque; the shunt motor is therefore less able to withstand overload than the series motor.

Fig. 29. The main-field flux Φ_{rs} as a function of the energizing current I_s (the 'open-circuit characteristic' of a d.c. motor). The magnetic saturation of the stator iron for increasing energizing current affects the shape of this characteristic.

Shunt and series motors behave quite differently when the load is varied. When the load is increased, the speed of a *shunt* motor falls only slightly. The extra power then delivered requires a proportional increase in the rotor current. The heavy portion of the curve *Sh* in fig. 28 indicates how the operating point moves in this case. The heavy portion of the curve *Se* indicates the shift of the operating point for a *series* motor when its rotor current increases by the same factor. This increase in the rotor current also results in a higher value for the main field, which contributes to an increase in the delivered torque; if the torque is doubled the rotor current only increases by about 40%, as against 100% for the shunt motor. This means that the series motor is more able to withstand overload than the shunt motor. Since the applied electrical power increases less for a series motor, the greater torque required can only be supplied at a significantly lower speed. A particular point to note is that if the loading torque is removed the motor speed rises sharply and will generally 'run away' to a speed above the maximum rated value.

If the series motor is supplied with a.c. current it gives a pulsating torque; with the field and rotor windings connected in series the main field and the rotor current are always in phase, so that their product al-

ways has the same sign. It is more difficult in this case to achieve good commutation, since the time-varying main field induces an additional voltage in the rotor coils. The stator core, and not just the rotor, also has to be laminated to prevent eddy-current losses. A series motor designed for use on either d.c. or a.c. current is known as a universal motor. Such motors are widely used in domestic equipment.

Brushless d.c. motor

A mechanical commutator requires periodic maintenance, is subject to wear, is noisy in operation and causes vibration and electrical interference. In many applications these effects are a nuisance or prevent the use of a mechanical commutator. For these reasons brushless d.c. motors have been under development for several years [*]. In these motors the commutation is obtained with electronic devices.

The rotor of a small brushless d.c. motor consists of a permanent magnet, while the stator has a multiphase winding. The windings can be wound on salient poles or in slots in the stator circumference. In small motors that have to provide a pulsation-free torque, i.e. no 'cogging', the windings are attached to a smooth stator bore in the form of 'air-gap coils'. *Fig. 30* gives a
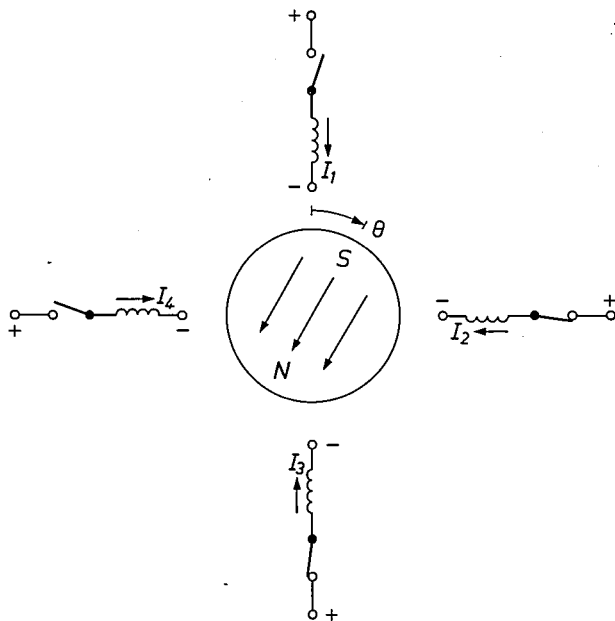
Fig. 30. Brushless d.c. motor. Opening and closing the switches of the four stator coils in succession produces a rotating stator field, which pulls the permanent-magnet rotor round with it. In practice the switches are transistors or thyristors. The control signals for these devices are derived from the position of the rotor, which can be monitored optically with photodiodes or magnetically with Hall elements.

[*] We intend to include an article on brushless d.c. motors in the second issue on electric motors. (Ed.)

schematic diagram of a brushless d.c. motor with a two-pole rotor and four stator windings; this arrangement of stator and rotor shows considerable similarity to a four-phase synchronous motor with a permanent-magnet rotor. However, in the motor of fig. 30 the four stator windings are supplied from a d.c. source, in a way that recalls the supply arrangements for the stepping motor. The notable difference, however, is that the opening and closing of the switches between the windings and the source are determined only by the position of the rotor and the desired direction of rotation. This is shown schematically in *fig. 31* for the case of clockwise rotation. Apart from brief transient effects, only two of the four windings are in use at any one moment. In small motors the switches consist of transistors, in large motors the rather less easily quenched thyristors are used.

There are various types of brushless d.c. motor, which can be classified by the nature of the elements for detecting the position of the rotor. First of all there are the Hall elements or field-dependent resistors. These are located in the rotor field, which varies with the position of the rotor [6]. The position of the rotor can also be established with the aid of photoelectric devices (photocells), magnetic [7] or electrostatic sensing elements. The electronic circuit is usually provided with a speed control and current limiting, giving a very adaptable d.c. motor that is particularly suitable for control purposes. Brushless motors are very suitable for applications where torque has to be uniform and
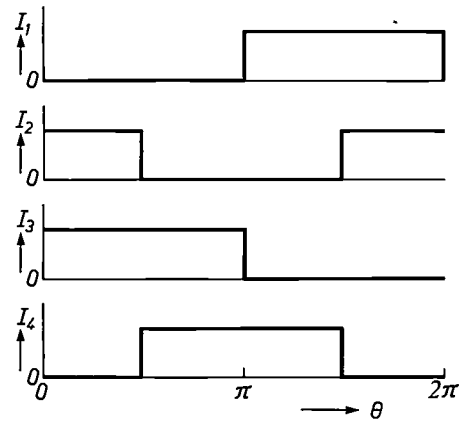
Fig. 31. Schematic representation of the cyclical variation of the four stator currents I_1 to I_4 for the brushless d.c. motor shown in fig. 30.

efficiency high; they can be used for the drive in battery-operated tape recorders and record players without the need for an intermediate transmission system.

Summary. The most important types of electric motor are described and the main characteristics are derived. Electric motors can be classified as a.c. or d.c. motors. A.C. motors are either synchronous or asynchronous. In synchronous motors the number of revolutions per second is equal to the frequency of the supply divided by the number of pairs of poles. The stator windings maintain a rotating stator field, which the rotor follows; the rotor can be either a permanent-magnet rotor or made from a material of high permeability (and non-cylindrical) in the reluctance motor, or of material with considerable hysteresis (hysteresis motor). In the asynchronous or induction motor the rotor has closed current circuits in which currents are induced; the rotor 'slips' with respect to the rotating stator field. D.C. motors have a commutator, which reverses the current in the windings after every half-revolution. The commutator usually consists of brushes and sliding-contact segments, but electronic commutation is sometimes employed (brushless d.c. motor). If the commutator motor is series wound it can also be run on a.c. current (universal motor).

[6] See: G. Bosch and J. H. H. Janssen, Integrated circuit with Hall device for brushless d.c. motors, Philips tech. Rev. 31, 366-368, 1970.
[7] W. Radziwill, A highly efficient small brushless d.c. motor, Philips tech. Rev. 30, 7-12, 1969.

Stability of small single-phase synchronous motors

H. Schemmann

Introduction

Electrical equipment nowadays does much of the work in and about the house that was formerly done manually. Electric power is used on a large scale for washing machines, vacuum cleaners and shavers, and also for tape recorders and record players. This has all become possible because such household machines have certain features in common: they are all simple to use, cheap to produce, use little power and are relatively cheap to service. In almost all cases the devices are required to operate from the ordinary single-phase mains supply.

For household use in the broadest sense, electrical energy is transformed into mechanical energy mainly by single-phase electric motors. Well known types of motors are synchronous and induction motors with their starting circuits, and hysteresis and series motors [1]. Their reliability, cost and efficiency must of course match the requirements of the devices in which they are used.

The development of oriented permanent-magnet materials with a high coercivity, such as ferroxdure, has led to the suggestion that synchronous motors with permanent-magnet rotors might offer advantages with regard to design, characteristics or efficiency. Such motors have no need for a commutator (which is subject to wear), they cause no electrical interference and they have a constant speed. Moreover, their efficiency can be high and the motor can be constructed from a small number of relatively simple components. *Fig. 1* shows the principle of such a motor [2]. A laminated stator is provided with a coil that can be directly connected to the mains. A permanently magnetized rotor of ferroxdure with two diametrically opposed poles is mounted on a shaft in the stator field and can be used to drive some external load.

When the voltage is switched on, the motor should be self-starting and it should then rotate steadily in the same direction. Our permanent-magnet synchronous motors, like synchronous motors in general, do not meet these requirements unless special measures are taken. A well known and generally used method of starting a single-phase synchronous motor is the use of an auxiliary coil to give a rotating field, in combina-

tion with an induction cage or a hysteresis cylinder on the rotor. In this way the motor can start to rotate asynchronously in one direction. This solution, however, requires modifications to both the rotor and the stator and should therefore be avoided if possible.

Another difficulty in purely single-phase motors is that while the stator field does in fact change sign it has a fixed direction: it does not rotate, as in rotating-field motors. The direction of rotation of such a single-phase motor is therefore indeterminate.
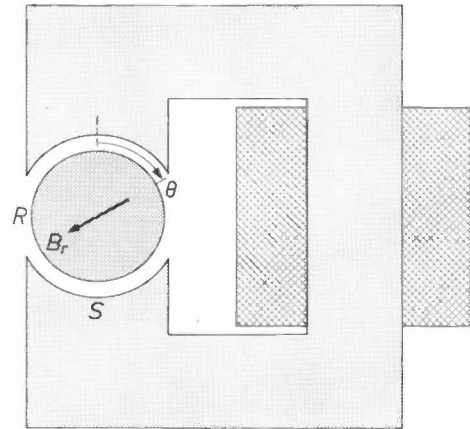
Fig. 1. Principle of a single-phase synchronous motor with a permanent-magnet rotor. R rotor, magnetized in the direction B_r. S stator. θ angular coordinate defining the rotor position.

Furthermore, a synchronous motor can behave as an oscillating system because there is an elastic coupling between the rotor and the rotating stator field. If the damping in this system is too small, oscillations can be set up that allow the motor to fall out of step with the field. With a non-conducting rotor, as in the present discussion, there is no electrical damping in the rotor at all. The stability of the rotary motion therefore requires special attention in such motors.

The advantages mentioned above were however sufficient to indicate a further study of these problems. In this article particular attention will be paid to the starting and stability of the motor.

Dr H. Schemmann is with Philips Forschungslaboratorium Aachen GmbH, Aachen.

[1] E. M. H. Kamerbeek, Electric motors, this issue, p. 215.
[2] R. Thees, Philips tech. Rev. **26**, 143, 1965 and Elektrotechn. Z. A **87**, 171, 1966.

Starting

Normal polyphase synchronous motors with no damping or starting cage will not in general start to rotate unaided when the stator winding of the stationary motor is connected to the supply. The polyphase supply produces a synchronously rotating field immediately but owing to its inertia the rotor cannot usually follow this field [3]. There are now two ways of remedying this situation. The motor can be brought up to speed mechanically from outside or the frequency of the supply voltage can be reduced until the rotor is in step with the field; the frequency is then increased slowly until the rotor reaches the desired speed.

In single-phase synchronous motors an auxiliary field is usually excited, as mentioned above, which combines with the main field to give a rotating magnetic field. This can be produced by means of an auxiliary winding on the stator, connected to the supply via a capacitor, or by including one or two short-circuited turns around part of the stator poles (shaded pole) [4]. Electric motors with a large moment of inertia or high speed usually run asynchronously during starting, that is to say, starting takes place by virtue of currents induced in the rotor by the stator field. Starting of hysteresis motors will not be discussed here.

Single-phase synchronous motors with a non-conducting permanent-magnet rotor cannot be started by any of the methods just mentioned. When the supply is switched on the rotor is subject only to an alternating field that only changes sign; it does not change direction. This field exerts a torque on the rotor only when the rotor field and the stator field do not lie exactly in the same direction. Their directions coincide twice per revolution; the rotor must be prevented from coming to rest in either of these two positions since it would then remain there.

The rotor of the motor shown in fig. 1 is subject not only to a torque exerted by the stator field but also to a magnetic torque (the reluctance torque) even when there is no current in the stator winding. This is because the total magnetic energy of the system depends on the orientation of the rotor magnet between the stator poles. This reluctance torque can be used to prevent the rotor from coming to rest with its field parallel to the stator field. This is just what does happen in a symmetrical configuration like that of fig. 1. If however the space between the stator poles is made asymmetrical so that the air gaps between rotor and stator along one diameter are smaller than those along another diameter (fig. 2) then the rotor will assume a rest position in which the magnetic resistance is at a minimum. The stator field is hardly affected by the above asymmetry: to the stator the rotor appears to be part of the air gap because of the relatively low permeability of the rotor

material. If $\hat{T}_0$ is the amplitude of the reluctance torque and γ the angle between the rest position (determined by the reluctance torque and the direction of the stator field [5] then the torque acting on the rotor in the undesired 'parallel-field' position is $T_0 = \hat{T}_0 \sin 2\gamma$. If the rotor is in fact to be moved from this undesired position, then T_0 must be larger than the total frictional torque T_{fr} acting on the shaft of the rotor.

In the electric motors considered here, the critical frictional torque is much smaller than the maximum torque of the motor. These motors are therefore only suitable in applications where the external frictional torque at the instant of starting is not too large.

Now let us consider the starting process in more detail. The behaviour encountered here is different from
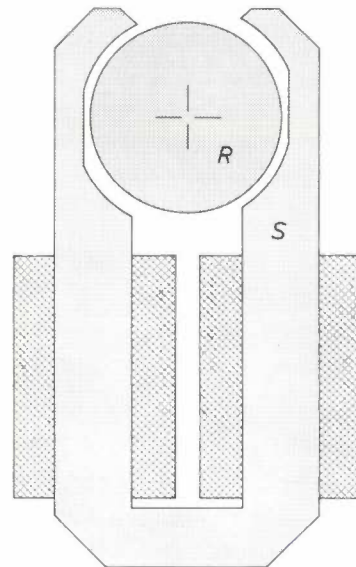
Fig. 2. Configuration of a single-phase motor with asymmetric stator poles. A motor like this is self-starting. R rotor. S stator.

the well known starting effects in asynchronous or d.c. motors; in these motors the rotor begins to rotate in one particular direction and then rotates faster and faster in that direction. In the single-phase synchronous motors considered here there is in general a complicated aperiodic transient behaviour. As *fig. 3* shows, the angular velocity is alternately positive and negative: the motor thus changes its direction continually. This irregular movement continues until the variables of the motion (speed, rotor position with respect to stator field, stator current and phase angle of this current) have adjusted themselves to the values corresponding to uniform rotation. The final direction of rotation depends on the initial conditions: in a given situation the phase of the voltage at the instant of switching on is particularly important.
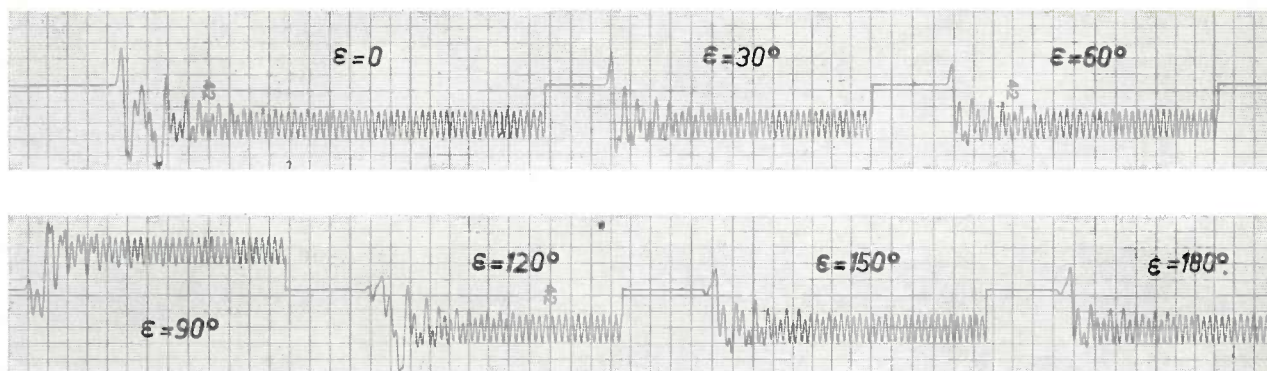
Fig. 3. Starting behaviour of a single-phase synchronous motor with a parmenent-magnet rotor. The angular velocity is plotted as a function of time for various values of the phase angle ε of the supply voltage at the instant of switching on. The starting behaviour and the direction of rotation are affected by this phase angle but the motional state finally achieved is always the same.

A motor behaves in the manner shown in fig. 3 only when the moment of inertia of the rotor and the load attached to it is not too great. If the combined moment of inertia is too great the rotor will only oscillate about an equilibrium position. In general it is not possible to start a single-phase synchronous motor by a gradual increase in the frequency of the supply voltage.

This rather summary discussion of methods for self-starting shows that a single-phase synchronous motor with a permanent magnet rotor can only be used to drive devices with a low static friction or which are coupled to the motor only after it has started. Moreover, the device to be driven must have a small moment of inertia and the direction of rotation must be unimportant.

The limitations imposed by the requirements of a small frictional torque and a low moment of inertia can largely be overcome by the use of a mechanical coupling that reduces the mechanical load on the motor to an acceptable value during starting. A number of schemes have been investigated, all depending on frictional effects. In the design of such couplings it is especially important that there should be no slip in the coupling after starting, for this would mean that the mechanical power of the motor could not be fully used.

If necessary the direction of rotation can be predetermined by means of a mechanical device that prevents rotation in one direction. In each case these extra complications must be weighed against the likely advantages.

Experimental investigation of the motor behaviour

We shall now examine the behaviour of a single-phase synchronous motor with a permanent-magnet rotor that has run up to speed after switching on and does not, as described above, remain stationary or just perform small-amplitude oscillations. An electric motor is only suitable for practical applications if once it has started it continues to rotate in the same direction with little variation in speed. It can be shown that to achieve this, the parameters of the motor, the supply voltage and the load must be matched to each other between fairly narrow limits. If this is not the case, a *perturbed* motion will ensue that differs considerably from the desired *unperturbed* motion [6]. A good overall picture of the possible rotor motions can be derived from measurements of angular velocity as a function of time for various amplitudes of the supply voltage. *Table I* (p. 239) shows the results of such measurements on a motor designed for a voltage of 220 V, which therefore rotates with an unperturbed motion at this voltage. There are, however, at this voltage speed variations at twice the mains frequency. This modulation of the rotational speed is due to periodic variations in the electromagnetic torque. In a purely single-phase synchronous motor, there is a field that rotates in the opposite direction to the rotor at twice the mains frequency and thus gives rise to an oscillating torque at this frequency [7]. For a motor of other dimensions or with another load, however, the various perturbed motions may take place at the nominal voltage.

The angular velocity is measured by means of a unipolar generator, which has negligible effect on the behaviour of the motor. In the diagrams in Table I the

[3] See for example R. Richter, Elektrische Maschinen, part I, 3rd edition, Birkhäuser Verlag, Basel 1967, p. 63.

[4] Further details are given on p. 229 of the article by E. M. H. Kamerbeek [1].

[5] The angle γ is the difference between the angles α and β defined by Kamerbeek [1].

[6] Definitions of the terms used here are given in J. G. Malkin, Theorie der Stabilität einer Bewegung, Oldenbourg, München 1959, p. 3.

[7] See p. 220 of the article by Kamerbeek [1].

angular velocity is plotted as a function of time. The graphs refer to the motion after starting transients have died away.

The phase of the voltage at the instant of switching on usually has no effect on the final rotation of the rotor; the situations 6 and 7 in Table I are exceptions.

The experiments show that the single-phase synchronous motor with the permanent-magnet rotor described has a useful motion only within a limited voltage range or, more correctly, in a limited range of motor parameters. The extent of this range and its limits are of course of interest. It should be remembered here that the motors must retain their unperturbed motion in spite of the spread in properties that normally occurs in industrial materials or as a result of the production process. The motors must also be able to take a load. Another interesting point is the extent to which the behaviour of the motor may be affected by nonlinear effects such as the saturation of the stator iron.

An attempt will be made to examine these questions using the equations of motion discussed in the next section, which describe the dynamic behaviour of the motor.

Equations of motion

A theoretical model of the situation must always satisfy two conditions. Firstly, it should allow the various relevant characteristics to be represented with sufficient accuracy. Secondly, the model must give a clear physical picture and it must be easy to manipulate. This second requirement means that simplifications have to be introduced and certain things neglected, in the hope, however, that nothing essential is lost. In the model of our motor it has been assumed that the magnetic state of the motor is uniquely determined by the stator current and the rotor position, in other words that hysteresis effects are completely neglected. Furthermore, saturation effects and eddy currents in the iron of the stator are neglected and the friction is assumed to depend linearly on the normal forces and to be independent of velocity (Coulomb friction). Finally it is assumed that the magnetic flux and the reluctance torque vary sinusoidally with the rotor position.

On the basis of these assumptions it is not difficult to write down the equations of motion. The first equation states that the sum of the voltages across the stator winding must be equal to the mains voltage. The sum is made up of the iR drop due to the current i through the coil of resistance R, the voltage $L\mathrm{d}i/\mathrm{d}t$ due to inductance L of the coil and finally the voltage $\mathrm{d}\Phi_{\mathrm{sr}}/\mathrm{d}t$ induced in the stator coil as a result of the change in

the flux Φ_{sr} of the rotor linked by the stator coil. The latter contribution is proportional to the angular velocity $\dot{\theta}$ of the rotor and is a sinusoidal function of the rotor position. The equation for the voltage across the stator coil is therefore:

$$\hat{v}\sin(\omega_{\mathrm{m}}t + \varepsilon) = iR + L\frac{\mathrm{d}i}{\mathrm{d}t} + \Phi_{\mathrm{sr}}\dot{\theta}\sin\theta,$$

where ω_{m} is the angular frequency of the mains voltage and ε its phase angle.

The second equation for the motor behaviour comes from the condition that the sum of the torques acting on the rotor must be zero. Apart from the frictional torque T_{fr} and the reluctance torque T_0 there is also an electromagnetic torque acting on the rotor, which is proportional to the stator current and to the rotor flux. The values of these torques must together be equal to the rate of change $I\ddot{\theta}$ of the angular momentum:

$$i\Phi_{\mathrm{sr}}\sin\theta - T_{\mathrm{fr}} - T_0 = I\ddot{\theta}.$$

The initial conditions of this equation are that at the instant the voltage is switched on ($t = 0$) the current is zero and the rotor is in the position $\theta = \gamma$, i.e. the position for static equilibrium.

In spite of the relatively simple form of the equations there is no general method for solving them. This is because of the nonlinearities that occur in the equations. (In the first equation the product $\dot{\theta}\sin\theta$, in the second $i\sin\theta$.) Attempts to linearize the equations yield no clear picture of the real situation; indeed the observed instabilities are probably directly due to these nonlinearities.

The equations have therefore been solved by means of an analog computer. This method gives a direct physical picture of the behaviour of the model. A more accurate solution on a digital computer is less attractive since programs for the solution of nonlinear differential equations require a great deal of computing time.

In the analog calculation a model with the same differential equations as the electric motor under investigation is set up in the computer. The operations such as summation, integration and multiplication are carried out by operational amplifiers. Important quantities can be read off during the calculations from instruments connected at appropriate points in the circuit [8].

The form of the theoretical model is determined by the equations of motion. To simulate a given motor, the various parameters of the motor must be determined and from these the coefficients of the equations must be derived. It should be noted that the inductance

[8] A description of the analog computer used is given in: H. Schemmann, Thesis, Eindhoven 1971, also published as Philips Res. Repts. Suppl. 1971, No. 5 (in German).

Table I. Typical curves for the motion of a single-phase synchronous motor with a permanent-magnet rotor, arranged in increasing values of the supply voltage.
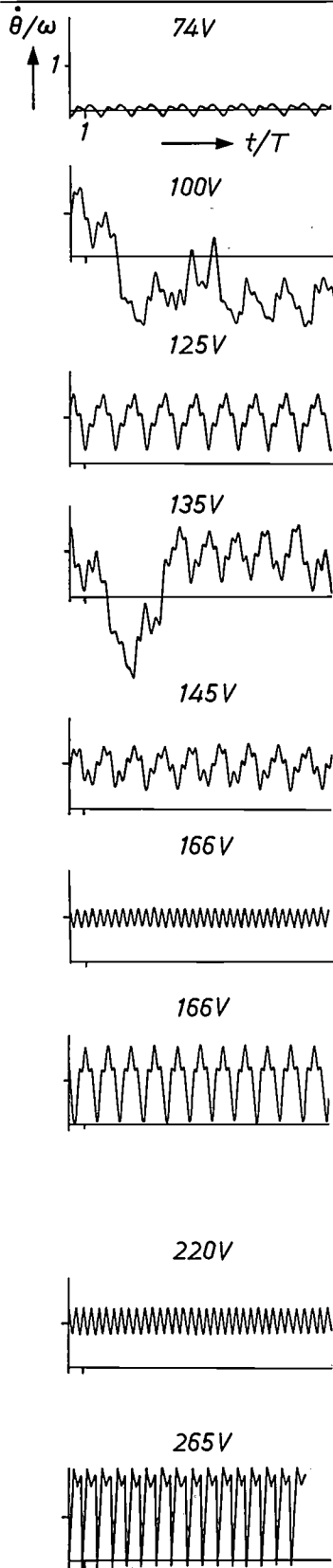
1. At low voltages the motor does not start. The rotor vibrates about the equilibrium position defined by the reluctance torque. (The same situation arises at the nominal voltage if the moment of inertia is too large.)

2. The rotor rotates but the motion is very irregular. The motion is not unidirectional even though the synchronous angular velocity is sometimes momentarily exceeded. The rotor changes its direction of rotation in a quite irregular manner.

3. The rotation has now become unidirectional. However, there are quite large periodic variations superimposed on the synchronous speed. Stroboscopic observations show that a steady-state motion is achieved with a repetition period of two revolutions of the rotor (25 Hz perturbation).

4. As the voltage is increased further the irregularities of the motion do not decrease. On the contrary, large new irregular and aperiodic motions occur in which the rotor again keeps reversing its direction of rotation. Apart from the voltage at which it occurs, this situation is no different from that described under 2.

5. When the voltage is increased still further, the motion becomes unidirectional again. Strong periodic variations are still present (hunting) but the motor runs with a mean speed equal to the synchronous value.

6. The amplitude of the hunting now becomes smaller; above a particular voltage the oscillations disappear entirely. For the first time the motor assumes its state of unperturbed rotation as at the nominal voltage (see 8).

7. At the same voltage at which this unperturbed motion occurs, however, the rotor can also run with very large speed variations. The rotor periodically comes almost to rest; the mean rotor speed is, however, the synchronous value. These variations are repeated every $1\frac{1}{2}$ periods ($33\frac{1}{3}$ Hz perturbation). Which of the two states of motion, 6 or 7, takes place is determined by the phase of the supply voltage at the instant the motor is switched on.

8. At the nominal voltage the rotation is again unperturbed. The angular velocity is now modulated only at twice the mains frequency. The modulation in speed is 20-40 % of the mean synchronous value. If a large moment of inertia is coupled to the motor, the modulation depth becomes smaller. The unperturbed motion is stable for voltage variations that are not too large, and the motor can drive a load.

9. Increasing the voltage above the nominal value again gives rise to large variations in the angular velocity. The rotor comes to rest once per revolution and the direction of motion may even be momentarily reversed. Within milliseconds the rotor then assumes velocities far above the synchronous value and is then immediately slowed down (50 Hz perturbation).

   If still higher voltages are applied, the accelerating torques are so large that any regular motion is quite impossible.

of the stator coil(s) and the value of the rotor flux must be determined under conditions that approach the actual working conditions as nearly as possible. This is necessary because in the equations of motion a linear relation has been assumed between stator current and stator field and in practice this is not the case.

very good. The 50 Hz perturbation, the unperturbed range and also the $33\frac{1}{3}$ Hz perturbation and its adjacent transition region occur in the model at almost the same voltages at which they are observed in practice. At voltages less than 160 V it is noticeable that the model exhibits less tendency to oscillation than is observed in
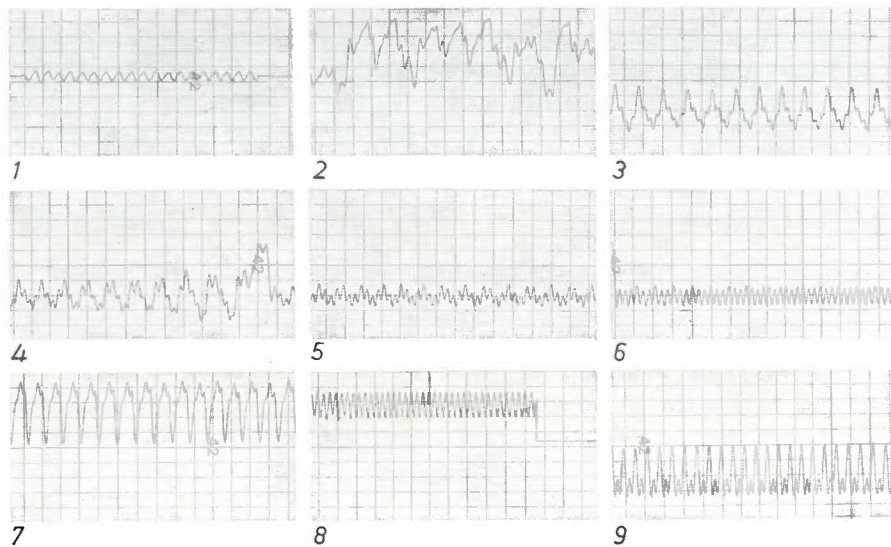
Fig. 4. Rotor motion after starting transients have died away, determined by an analog computer. The curves *1*, ... *9* show the behaviour as the supply voltage is increased; curve *8* refers to the nominal voltage. The types of motion agree qualitatively very well with the results measured on a practical motor, as given in Table I.

Before conclusions can be drawn from the model, it must be established that, in essentials, the model describes the real situation sufficiently well. This point has been investigated by determining how the model describes the angular velocity as a function of time after switching on the voltage, as was done earlier for the motor itself.

The results are shown in *fig. 4*. Comparison of these results with the results of the measurements on the motor, as given in Table I, show that the equations represent a behaviour completely analogous to that of the motor. Not only do the equations reproduce the details of the motion observed in the motor but there are no extraneous effects, i.e. effects not observed in the motor. This implies that neither the perturbed motion nor the unperturbed motion are caused by effects that have been neglected in the formulation of the model.

If the voltages corresponding to the limits between the regions of perturbed and unperturbed motion are compared, the following conclusions can be drawn (see *fig. 5*). Particularly at voltages above 160 V, which is the range of most interest, the agreement between the calculated behaviour and the observed behaviour is

practice. This difference is connected with the fact that the curvature was neglected in the magnetization characteristic of the stator iron. Since voltages below 160 V are of little practical importance, it is not necessary to revise the model in this connection. Apart from this the choice of the model is seen to be justified by the results.

Periodic solutions of the equations of motion

We have seen that the motor can rotate with a constant mean speed. Superimposed on this mean value are large periodic variations with frequencies that are simply related to the supply frequency.

In general a synchronous motor, once started, must run at a constant mean speed; for a rotor with two poles the mean speed must correspond to the frequency of the supply. This means that the mean of the speed variations over a long time must be zero, or

$$\frac{1}{\tau} \int_0^\tau \left(\frac{d\theta}{dt} - \omega_m\right) dt = 0, \qquad (1)$$

where the integration time τ extends over an integral

number of periods of the speed variations. For steady-state motion, i.e. if the speed variations are periodic, then these variations can be written as the sum of a number of simple harmonic motions:

$$\theta - \omega_{\mathrm{m}} = \sum_{\nu=1}^{n} A_\nu \sin (a_\nu \omega_{\mathrm{m}} t + \phi_\nu), \qquad (2)$$

where the factors a_ν are provisionally arbitrary real numbers. Inserting (2) in (1) gives:

$$\frac{1}{\tau} \int_0^\tau \sum_{\nu=1}^{n} A_\nu \sin (a_\nu \omega_{\mathrm{m}} t + \phi_\nu)\mathrm{d}t = 0 .$$

Either the integration or the summation may be performed first. The expression is identically zero, if the separate terms of the sum are each equal to zero:

$$\frac{1}{\tau} A_\nu \int_0^\tau \sin (a_\nu \omega_{\mathrm{m}} t + \phi_\nu)\mathrm{d}t = 0 . \qquad (3)$$

The instants in time $t = \tau, 2\tau, 3\tau, \ldots$, after which the rotor motion is repeated, must be synchronous with the supply frequency. In the stable unperturbed state of motion, the fundamental of the speed variations has a period equal to half the period T of the supply voltage. The integration time τ can thus be taken equal to $\tau_1 = T/2 = \pi/\omega_{\mathrm{m}}$. This corresponds to a frequency of twice that of the supply voltage; it depends on the mains-frequency component of the stator current.

Other possibilities of synchronization occur when the period over which speed variations average out is an integral multiple of τ_1; then $\tau = k\tau_1 = k\pi/\omega_{\mathrm{m}}$, where $k = 1, 2, 3, \ldots$ The rotor then runs synchronously with the voltage supply. The condition (3) can now be written, on omitting the factors before the integral, as:

$$\int_0^{k\pi/\omega_{\mathrm{m}}} \sin (a_\nu \omega_{\mathrm{m}} t + \phi_\nu)\mathrm{d}t = 0$$

or

$$\cos (a_\nu k\pi + \phi_\nu) = \cos \phi_\nu .$$

This condition is fulfilled when $a_\nu = 2l/k$, where $l = 1, 2, 3, \ldots$. It follows that steady-state motion is obtained when a_ν has the following values:

$k = 1$	$a_1 = 2, 4, 6, \ldots$
$k = 2$	$a_2 = 1, 2, 3, \ldots$
$k = 3$	$a_3 = \frac{2}{3}, \frac{4}{3}, \frac{6}{3}, \ldots$
$k = 4$	$a_4 = \frac{1}{2}, \frac{2}{2}, \frac{3}{2}, \ldots$
$k = 5$	$a_5 = \frac{2}{5}, \frac{4}{5}, \frac{6}{5}, \ldots$

For the various values of a_ν, the motional states are represented by the sum of a number of sine terms. We thus find, for $k = 1$:

$$\theta = \omega_{\mathrm{m}} + A_1 \sin (2\omega_{\mathrm{m}} t + \alpha_1) + A_2 \sin (4\omega_{\mathrm{m}} t + \alpha_2) + A_3 \sin (6\omega_{\mathrm{m}} t + \alpha_3) + \ldots,$$

THEORETICAL MODEL   PRACTICAL MOTOR
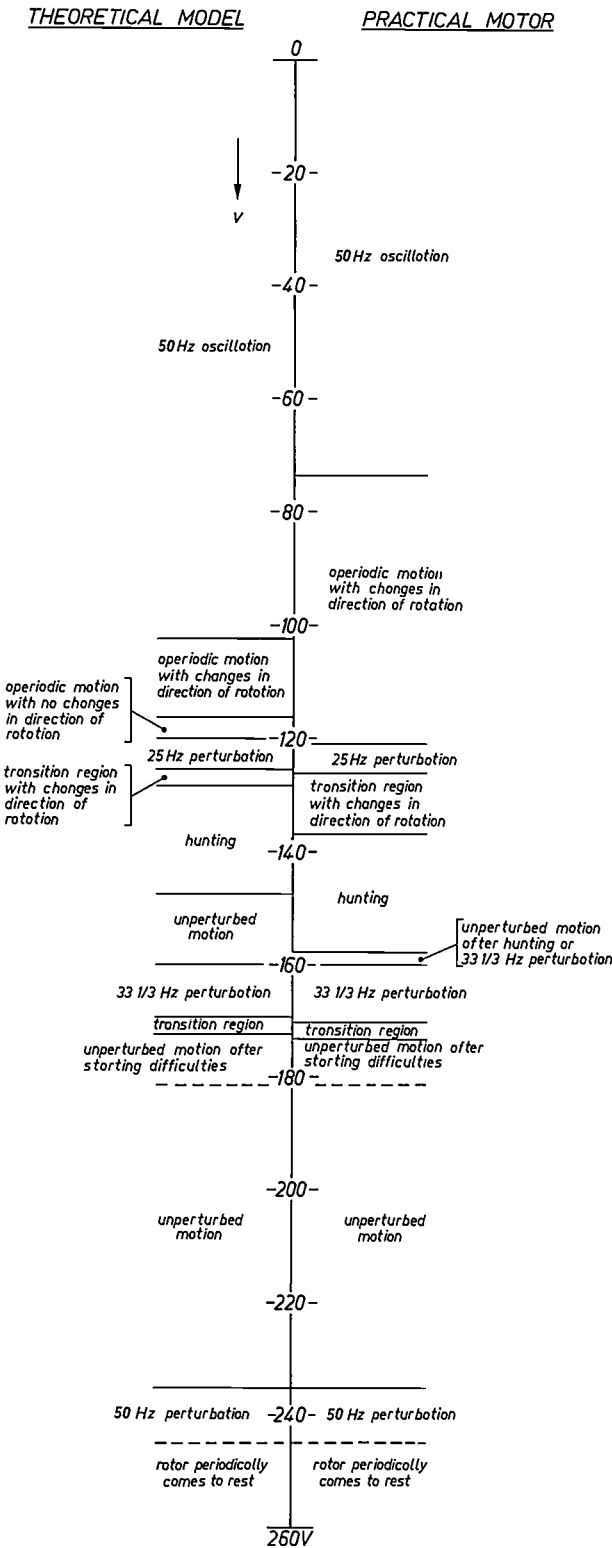
Fig. 5. The various regions of perturbed and unperturbed rotor motion. *Left:* results from the analog computer. *Right:* results measured on an actual motor.
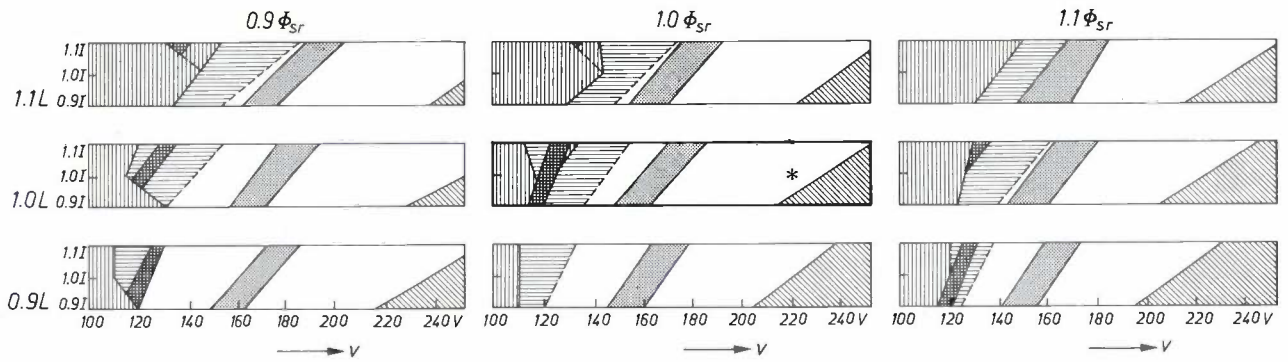
Fig. 6. Behaviour of the motor for a 10% variation in each of the following parameters: the rotor flux Φ_{sr} linked by the stator, the inductance L of the stator coil or coils and the moment of inertia I of the rotor. In each diagram the moment of inertia is plotted vertically and the supply voltage v horizontally. In the central diagram * indicates the situation for a motor operating correctly.

- ▥ aperiodic motion with changes in direction of rotation.
- ▤ the same, but direction of motion does not change
- ▦ 25 Hz perturbation
- ▩ 33⅓ Hz perturbation
- ▨ 50 Hz perturbation
- ☐ unperturbed motion

and, in accordance with the definition given above, this equation describes an unperturbed motion. Larger values of k give the various perturbations indicated above by the fundamental frequency of the appropriate variation. For a 50 Hz supply frequency there is therefore, for $k = 2$, a 50 Hz perturbation:

$$\theta = \omega_m + B_1 \sin(\omega_m t + \beta_1) + B_2 \sin(2\omega_m t + \beta_2) + B_3 \sin(3\omega_m t + \beta_3) + \ldots$$

and for $k = 3$, a 33⅓ Hz perturbation:

$$\theta = \omega_m + C_1 \sin(\tfrac{2}{3}\omega_m t + \gamma_1) + C_2 \sin(\tfrac{4}{3}\omega_m t + \gamma_2) + C_3 \sin(\tfrac{6}{3}\omega_m t + \gamma_3) + \ldots$$

Perturbations are also possible for larger values of k. The longer the period of a perturbation, the more complicated the resulting motion.

A periodically perturbed motion in which the accelerations are very large and the frequencies only slightly different from those mentioned here is not possible. The motor would then fall out of step because the angular positions of the rotor would exhibit steadily increasing deviations from the synchronous values.

The perturbed motions found can all be considered as subharmonic resonances of a periodic driving torque with a frequency equal to twice the supply frequency. Such a torque is always present in a single-phase motor, owing to the reverse field that rotates around the surface of the rotor at this frequency.

Stable regions

Experimental and theoretical investigations of the behaviour of our motor have yielded the following result. The perturbed motions that remain after the starting transients have died away are characteristic of the motor. The unperturbed motions can be regarded as transition states between the upper limit of the one perturbed region and the lower limit of the next. To be useful, an unperturbed region must be at least wide

enough to cover the spread in supply voltage that may be expected. There must be no chance that any perturbations should arise as a result of the spread in material properties, dimensions or load.

The effect of the spread in the various parameters can be very conveniently investigated with the analog computer. Using the model discussed earlier, it has been found that the inductance of the stator coils, the moment of inertia and the flux linkage of the rotor are particularly important in determining the behaviour. The effect of a 10% variation in these parameters is shown in the stability diagram of *fig. 6*. The following conclusions can be drawn from this diagram. Firstly, increasing the flux linkage of the rotor displaces the perturbed regions towards lower voltages. It can also be seen that an increase in the inductance of the stator coils results in a displacement of the perturbed regions to higher voltages. Finally, any increase in the moment of inertia also displaces the perturbed regions towards higher voltages. The value of the inductance of the stator coils also has a strong effect on the region mentioned earlier of oscillation below the 33⅓ Hz perturbation region. Unperturbed motion takes place only with small inductances. If the inductance increases, the unperturbed region becomes narrower, and eventually disappears altogether. Increasing the resistance of the stator coils has just the opposite effect.

It can be seen that changes in the parameter values only lead to displacement of the limits between the various stability and instability regions. No new effects are observed. In designing a motor for a given application care must be taken to obtain a sufficiently wide stability region around the nominal supply voltage and to ensure that this stability region continues to exist for all possible values of the mechanical load.

Both theory and experiment show that the single-phase synchronous motor with permanent-magnet rotor remains stable at loads equal to the maximum

theoretical value, provided that a small part of the load consists of a velocity-dependent damping. This is the case in many applications of these motors. A reduction of the Q (quality factor) $\omega_m L/R$ of the stator coil has, as in the unloaded oscillation of the rotor, a stabilizing effect. The motor can in any case be used for applications in which the load is of a frictional nature (i.e. both Coulomb and velocity-dependent friction).

Applications

The single-phase synchronous motor with permanent-magnet rotor as described is particularly suitable for driving small mass-produced articles such as domestic appliances. As we have seen, it is not a universally applicable motor characterized simply by voltage, power and speed. Because of problems with starting, rotation direction and the stability, the motor is only suitable as a drive for certain types of load. The motor therefore has to be specially designed for the load it must drive in practice and it is because of the development work required that it is only really suitable for mass-production applications.

For loads with a small moment of inertia and low bearing friction there are few difficulties — certainly not when the load is coupled to the motor after starting and when the direction of rotation is not important. The motor is therefore very suitable for driving a reciprocating device where the direction of rotation is immaterial. Moreover, the frictional load is then dependent on the position of the driven components; it is therefore possible to couple the motor and load in such a way that the friction is least in the most critical rotor position (the position with the rotor field parallel to the stator field). Also, with a reciprocating motion, the inertia of the reciprocating parts is much reduced in the transmission to the rotor. Compared with a vibrator drive, there is the advantage that the amplitude of the motion does not depend on the load or on the supply voltage; the motor can also be designed so that the whole device still functions properly at various supply frequencies. Furthermore, the loading of a reciprocating device has no effect on the speed of synchronous motors, unlike induction and series motors.

Fig. 7 shows an electric hair/massage brush in which the attractive features of the synchronous motor are put to use. The maximum available power developed by this motor, depicted in *fig. 8*, is about 7 W and the motor can be run from either a 220 V/50 Hz supply or a 120 V/60 Hz supply. The stator coils are encapsulated in plastic so that the motor is protected against moisture. A similar type of motor can be used for the drive of a trimmer.
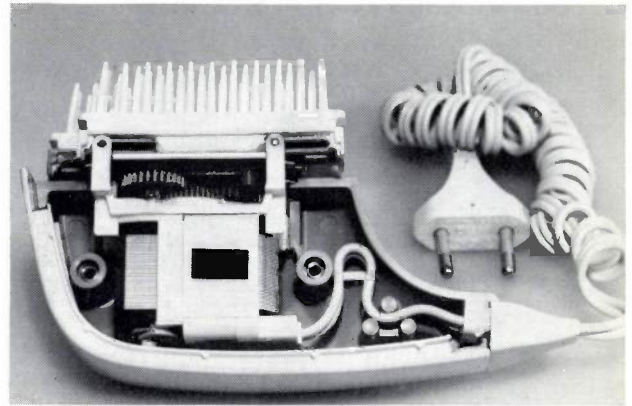
Fig. 7. Hair and massage brush driven by a single-phase synchronous motor with a permanent-magnet rotor. The mechanical coupling between the rotary motion and the reciprocating motion can be seen between the motor and the brush.
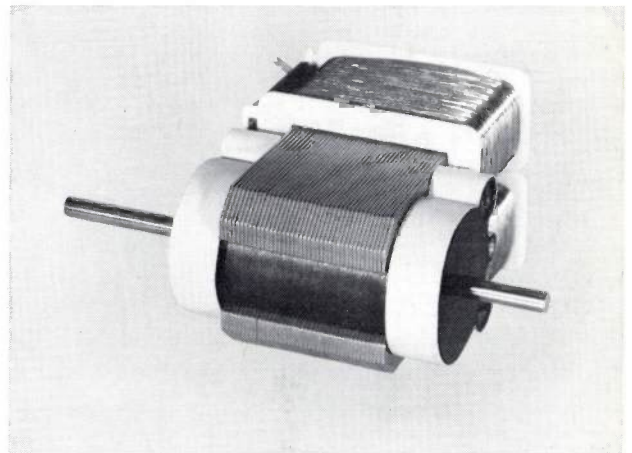
Fig. 8. The motor used in the brush of fig. 7.

For other applications a suitable coupling is required that restricts the moment of inertia during starting and ensures unidirectional rotation. With these measures, the possibility of applications in rotary shavers and similar devices can be considered.

Summary. The small single-phase motor described here with a permanent-magnet rotor is a small and efficient motor of very simple construction: a magnetized cylindrical rotor of ferroxdure and a laminated stator with a field winding. To make the motor self-starting without any special extra provisions (auxiliary coil with capacitor, or 'shaded poles') the air gap of the motor is made asymmetrical, so that the rotor field is not parallel to the stator field in the rest position. Once the motor has started, then depending on the motor parameters and the load regions of perturbation can affect the movement; the perturbations consist of very strong periodic variations in the speed of rotation. Studies made on the motor with an analog computer show that the existence of perturbed and unperturbed regions is an essential feature of this type of motor, which can be traced back to non-linearities in the equations of motion that describe the behaviour of the motor. It is shown that these perturbed movements also satisfy these simplified equations. It is relatively easy to design a motor in such a way that a sufficiently wide unperturbed region is obtained around the nominal values of the supply voltage and the load. Finally, a number of possible applications for the motor are described. Since a special version of the design has to be produced for each application the motor is only of interest for driving devices that are to be produced in quantity.

Moving-coil motors

J. H. M. Hofmeester and J. P. Koutstaal

Small d.c. motors are used in very large numbers in playback equipment, toys, measuring recorders, etc. These motors owe their wide applications to a number of features which can be made use of singly or in combination, depending on the particular requirements. For example, the efficiency is high, even for very small motors; the delivered torque is uniform, and there is ample scope for selection and control of the motor speed. Further features which may be of importance are the rapid start, the ease with which the direction of rotation can be reversed and the possibility of using a battery as power supply. The rotor of such motors often consists of a laminated ferromagnetic material, with the rotor windings wound round the stack of laminations. *Fig. 1* shows by way of example a simple version with three rotor lobes. The stator generally consists of a permanent magnet in the form of a hollow cylinder, which is sometimes divided into a number of segments.

Despite the utility of this type of motor for many applications, it has certain disadvantages when a very uniform torque is required, e.g. in tape recorders. Since there are slots in the rotor, it has a number of preferred positions, sometimes called 'reluctance positions'. The corresponding magnetic reluctance torques can be felt when the shaft of the motor is turned by hand. It will be clear that these reluctance torques give fluctuations in the torque delivered by the motor; this has an adverse effect on the quality of audio recording and playback. The reluctance torques can also cause motor noise, since there is always a little play in the bearings. The reluctance torques can be reduced by increasing the number of rotor lobes, but for given motor dimensions this can only be done at the expense of the efficiency, since the space factor of the winding is then decreased. Another solution is to minimize the width of the rotor slots, but this also can only be done within certain limits since the slot has to be wide enough for the winding process.

The impossibility of eliminating the reluctance torques from motors with an iron rotor while continuing to satisfy all the other requirements involved led us to design three d.c. motors with no iron in the rotor. Such motors, which are called moving-coil motors, offer the following advantages:

J. H. M. Hofmeester and J. P. Koutstaal are with the Philips Audio Division at the 'Johan de Witt' Works, Dordrecht, the Netherlands.

1. Reluctance torques cannot arise since there is no iron.

2. Since there are no discrete rotor lobes, the rotor can be wound uniformly, which makes it possible to divide the winding into a large number of coils. This also helps to reduce the fluctuations in the delivered motor torque, since a large number of coils with corresponding commutator segments reduces the torque pulses resulting from commutation. This feature makes this type of motor very suitable if a low and uniform speed of rotation is required.

3. The moment of inertia of the rotor can be made smaller than that of a rotor with an iron core in a comparable motor (same power consumption, same torque-speed characteristic).

4. Iron losses (hysteresis losses and eddy-current losses) cannot occur in the rotor. This is especially important in small motors, where the delivered motor torque can be of the same order of magnitude as the loss torque caused by an iron core (a situation which leads to a very low efficiency).

5. The inductance of the rotor is low, which improves the commutation: there is less electrical interference, and the life of the commutator (and hence of the motor) is increased.

We shall now discuss the basically simple principle of the d.c. moving-coil motor, with reference to *fig. 2*.
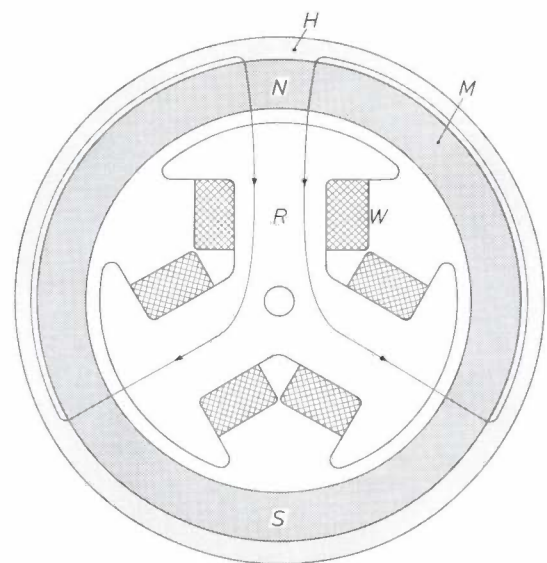
Fig. 1. Example of a d.c. motor with iron rotor. *H* housing. *M* cylindrical permanent magnet with poles *N* and *S*. *R* rotor, consisting of a stack of laminations. The winding *W* is situated in the slots between the rotor lobes (three in this example).

Fig. 2. Principle of the moving-coil motor. H housing. M permanent magnet (stator). W one turn of the coil (rotor). The side of the coil is of length l. The magnet M gives a magnetic flux density B in the air gap d. When a current i flows in this turn, a Lorentz force F is exerted on it. The radius of the winding is b.

The stator consists of a permanent magnet M, which produces a magnetic flux density B in the air gap d between the magnet and the housing H (which is made of a ferromagnetic material). The rotor winding W rotates round a vertical axis in this air gap (only one turn of the winding is shown). If the flux density B and the current i in the turn have the directions shown in the figure, the rotor will rotate anticlockwise under the influence of a Lorentz force F applied to one side of the turn; the magnitude of this force is liB, where l is the length of the side of the turn in question. The torque delivered per turn, T_e, is equal to $2Fb$ (b is the radius of the winding).

Construction

Fig. 3 shows a cross-sectional view of one of our moving-coil motors. The stator consists of a permanent magnet M mounted on the steel motor housing H by means of a support *7*. This steel housing is also a part of the magnetic circuit; the magnet has two diametrically opposite poles.

The rotor is formed by a winding W wound on a coil former *5* with cover *6*. The coil former is fixed on a shaft *8*. The winding consists of a large number of turns distributed uniformly round the circumference, and is divided into five or nine coils, depending on the type of motor.

The commutator *2* is built up of flat segments (see also fig. *4b*), one per rotor coil. These segments are gold-plated by a special technique.

The brush unit consists of two brush springs *1*, mounted in an injection-moulded plastic bridge *4*. The
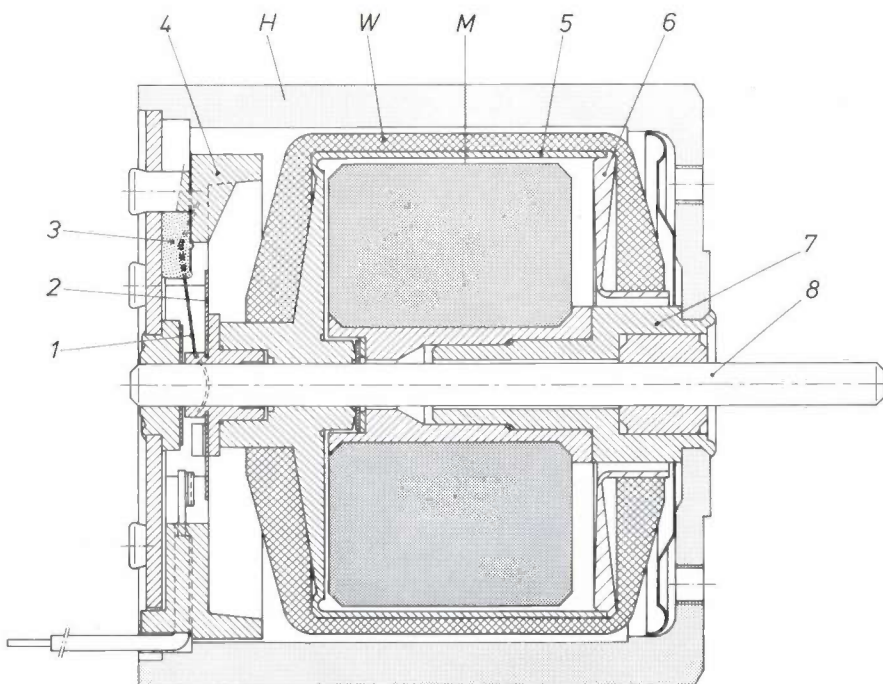
Fig. 3. Cross-sectional view of a practical version of the moving-coil motor. *1* brush spring. *2* commutator segments. *3* damping compound. *4* plastic bridge carrying the brush springs. H motor housing. W coil. M permanent magnet. *5* coil former with cover *6*. *7* support for mounting the magnet on the housing. *8* motor shaft.

brush springs are silver-plated at the place where they make contact with the commutator. Damping compound 3 is applied to damp the vibrations of the brush springs.

Winding and magnet

It can be seen from fig. 3 that the coil former has central projections at each end: these are necessary for the insertion of the magnet support and for attaching the coil former to the motor shaft. This arrangement does not permit purely diametral winding, and the winding method shown in *fig. 4a* has to be used, in which the plane of the winding is slightly skewed with respect to the centre-line of the coil former.

A regular distribution of the turns around the circumference is obtained by rotating the coil former through a small angle α after each turn of the coil has

centrated at a small cross-section through the winding. As a result, the total length of the winding in a motor of a given size can be made less than in a moving-coil motor. Since the motor losses are mainly the sum of copper-resistance losses and iron losses, the higher dissipation in the coil of the moving-coil motor could partially cancel out the advantage of the absence of iron losses.

This difficulty can be avoided by making the permanent magnet of material with very good magnetic properties. This makes it possible to obtain a motor with good characteristics (in particular, high efficiency and short starting time). In the moving-coil motors described here we used 'Ticonal' 550 [*], a magnet steel which is given an anisotropic structure by means of a special heat treatment; this material combines a reasonably high remanence with a high coercivity (see
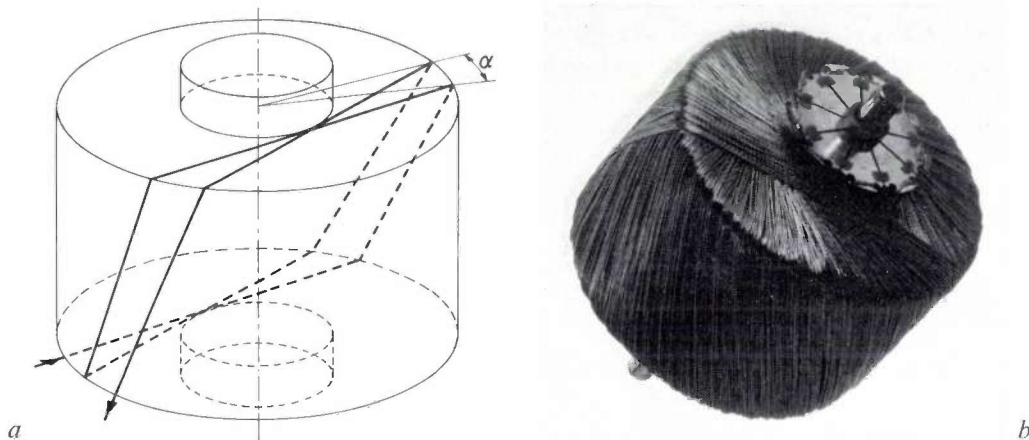
Fig. 4. *a*) Principle of the 'ball-winding' method which has to be used to leave the central space in the coil free. The turns are distributed over the circumference by rotating the coil former through an angle α after each turn. *b*) A rotor (coil + commutator) wound in this way.

been wound; the value of α is chosen so that the coil former will have rotated through exactly 360° after the desired number of turns. Because of the skewed position of the turns, successive turns will cross previously wound ones during the winding process. The coil obtained in this way looks rather like a ball of string and this winding method is therefore called 'ball winding'. A photograph of a wound rotor complete with commutator is shown in fig. 4b.

The air gap between the magnet and the housing is considerably wider than the thickness of the winding, since room must be left for the coil former and for clearance on each side. Since this space can only be provided at the expense of the magnetic flux, the moving-coil motor is at a disadvantage here compared with the iron-rotor motor (see fig. 1). Moreover, in the motor with iron rotor the magnetic flux can be con-

fig. 5). 'Ticonal' 550 has a $(BH)_{max}$ product of about 4.4×10^4 J/m³ (5.5×10^6 GsOe). In the operating range of the motors, the product BH lies between 4.4×10^4 and 3.5×10^4 J/m³. The high magnetic flux produced in the air gap as a result of the use of this material, and which is not of course limited in this type of motor by magnetic saturation of the core, compensates for the above-mentioned disadvantages. A further advantage is that with the geometry of coil and magnet used here, the ends of the coil are situated in the stray field of the magnet, so that the winding wire there does not merely increase the resistance of the coil, it also makes an appreciable contribution to the motor torque.

'Ticonal' 550 magnets are expensive compared with magnets made of the materials conventionally used in d.c. motors, such as ferroxdure. The winding method
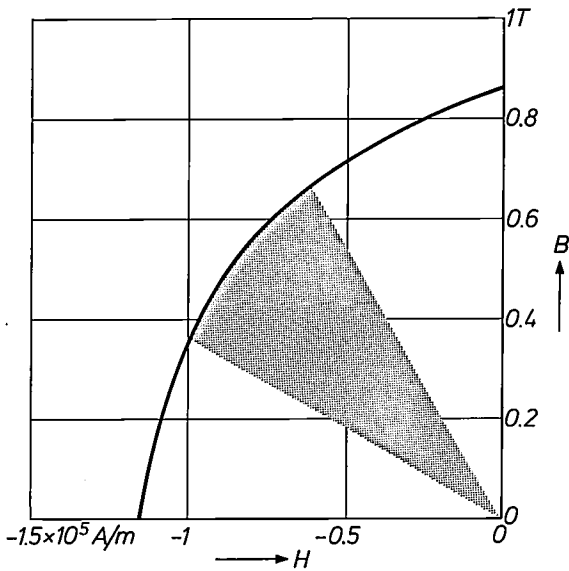
Fig. 5. Demagnetization curve of 'Ticonal' 550. B is the flux density in teslas (1 T = 10^4 Gs). H is the demagnetizing field-strength in A/m (1 A/m = $4\pi \times 10^{-3}$ Oe). The load lines for the magnets of the motors described in this article fall inside the shaded area.

used for the coil, and the complete construction of the motor are also more expensive. It is therefore likely that the application of the moving-coil motor will be restricted to cases where high-quality operation is required — in particular, uniform rotation, a low loss torque and a small moment of inertia. The last point is of importance for rapid motor starting, and rapid speed changing; the 'starting time constant' of the motor, which is a measure of its performance in this respect, should be kept low.

The starting time constant τ_s depends on a number of characteristic features of the motor. To illustrate the relation between this constant and the features in question, we shall derive an expression for τ_s. If we neglect the electrical time constant, equal to the ratio L/R of the inductance L and the resistance R of the motor, we can start by considering the linear relation between the motor torque T_e and the speed n found for a d.c. motor (fig. 6). This linear relation is represented by the equation

$$T_e = T_s(1 - n/n_0),$$

where τ_s is the starting torque and n_0 the speed at zero load. Mechanical and electrical losses are neglected here, which means
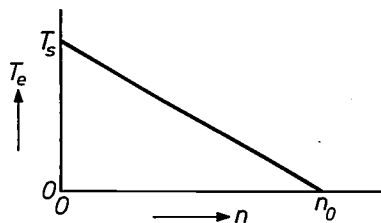
Fig. 6. Torque-speed characteristic of a d.c. motor. The motor torque T_e is plotted as a function of the speed n. T_s is the starting torque, and n_0 the zero-load speed.

that the motor approaches the zero-load speed n_0 asymptotically; at the speed n_0 it will rotate with a drive torque $T_e = 0$.

The time constant τ_s is defined for the *unloaded* motor, so that only the moment of inertia of the rotor is of significance. If we denote the latter quantity by I, the changes in speed are given as a function of time by:

$$T_e = 2\pi I \frac{dn}{dt}.$$

Combining the above two equations, we find that the speed a time t after switching on is given by:

$$n(t) = n_0(1 - e^{-t/\tau_s}),$$

where

$$\tau_s = 2\pi I n_0/T_s.$$

The starting time constant τ_s is thus proportional to the moment of inertia I, which can be made much smaller for an ironless rotor than for an iron rotor. If we want to derive the full advantage from this, we must ensure that the ratio n_0/T_s for the moving-coil motor is no larger than the value for an iron-rotor motor.

The zero-load speed is defined by:

$$V = 2\pi C n_0 \Phi_{rs},$$

where V is the applied voltage, C is a motor constant and Φ_{rs} the flux enclosed by the rotor coils [1]. The starting torque is given by:

$$T_s = C i_s \Phi_{rs},$$

where i_s is the current immediately after switching on. The current is therefore entirely determined by the motor resistance R and the applied voltage:

$$i_s = V/R,$$

so that

$$\frac{n_0}{T_s} = \frac{R}{2\pi C^2 \Phi_{rs}^2}.$$

The time constant is thus inversely proportional to the square of the enclosed flux. If we want to take full advantage of the small moment of inertia of the moving-coil motor, we must provide a powerful magnetic field. As mentioned above, we are not limited here by the magnetic saturation of the rotor core as in the case of a motor with iron rotor.

Practical realization

Fig. 7 gives the dimensions and other data for three of our moving-coil motors, together with a number of applications. Motor *A* was designed for use in a new version of the Pocket Memo dictating machine [2]. The good performance of the motor is reflected in the following four points:

1. Since the motor is small (the delivered torque is only 2×10^{-4} Nm), the low loss torque of the moving-coil motor (0.5×10^{-4} Nm) is particularly useful. As a result of this, the mechanical efficiency is as high as 80%, while the overall efficiency (54%) is also relatively high.

[1] See E. M. H. Kamerbeek, Electric motors; this issue, p. 215.
[2] An earlier version of the Pocket Memo is described in P. van der Lely and G. Missriegler, Audio tape cassettes, Philips tech. Rev. 31, 77-92, 1970.
[*] 'Ticonal' is a registered trade mark of N.V. Philips' Gloeilampenfabrieken.

The current required for driving this machine is only 20 mA.

2. The Pocket Memo has no capstan and hence no flywheel, so that a motor with a uniform speed is vital here to avoid wow and flutter.

3. The motor noise must be kept to a minimum, since the built-in microphone is very close to the motor. An important source of noise — bearing chatter due to the reluctance torques — is completely absent in the moving-coil motor.

4. As the motor is also very close to other electrical

The more powerful motor C is used in tape recorders with separate reels and in professional equipment such as measuring recorders and computers. The application in tape recorders is again based on the uniform rotation, even at low speeds where the effect of the flywheel is slight. The small time constant ($\tau_s = 19$ ms) is a great advantage for professional equipment.

These examples show that a motor with a moving-coil rotor may be the best choice in cases where uniform rotation, minimum motor noise and a small starting time constant are of importance.

	A	B	C
T_e	2×10^{-4} Nm	5×10^{-3} Nm	10^{-2} Nm
efficiency	54%	64%	82%
τ_s	56 ms	23 ms	19 ms
	5 rotor coils	9 rotor coils	9 rotor coils
	pocket dictating machines cameras	digital cassette recorders audio cassette recorders	tape recorders professional equipment, such as measuring recorders and computers
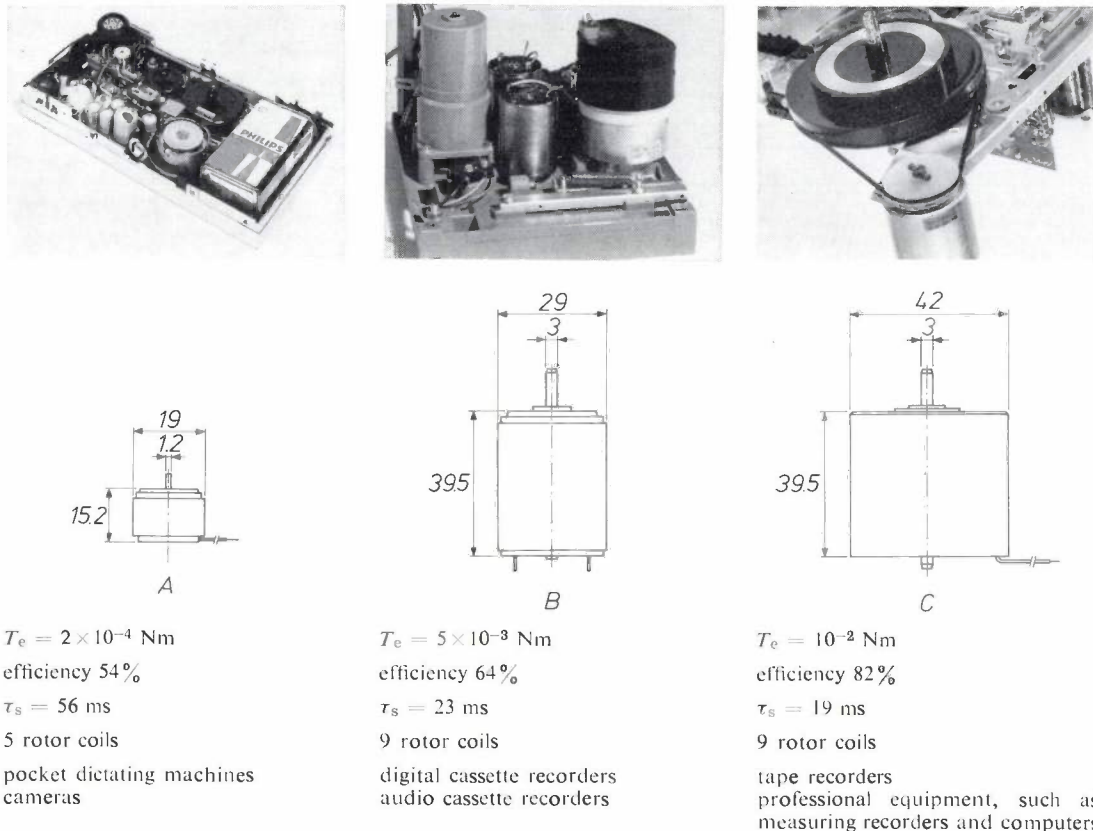
Fig. 7. Illustrations and data for three moving-coil motors. The torque T_e is given for a speed of 3000 rpm. τ_s is the starting time constant. The photograph for motor A shows this motor used in a Philips Pocket Memo dictating machine. The photo for motor B shows a model of a digital cassette recorder developed at Philips Electrologica. Apart from two type B motors, which are used for tape transport in both directions, this photograph shows (on the right) a type C motor with a tachosystem designed for tape-speed control. The photograph for motor C shows this motor mounted in a Philips tape recorder.

components in the dictating machine, the low rotor inductance — and therefore the low electrical interference level — is a useful feature.

Motor B was developed for cassette recorders (for both audio and digital data). The diameter is chosen so that four motors can be mounted under one cassette if required, thus permitting both cassette reels and both capstans to be driven directly (for tape transport in both directions). In tape recorders the direct drive is an added advantage; the moving-coil motor does not produce wow and there is no chance of wow arising from gearing. The small starting-time constant ($\tau_s = 23$ ms) is important for digital applications.

Summary. Moving-coil motors contain a rotor consisting merely of a coil with commutator rotating about a permanent magnet (the stator); conventional small d.c. motors have rotor coils wound on an iron core. The advantages of the ironless rotor include the absence of reluctance torques (so that it is easier to obtain uniform, low-noise motion) and the small moment of inertia, permitting short start and stop times. Moving-coil motors also have a high efficiency (particularly in the low-power models) because there are no iron losses. However, these motors have the disadvantage that the flux density inside the coil is lower than in an iron-rotor motor of comparable dimensions. In the motors described in this article, this disadvantage is minimized by making use of the permanent-magnet material 'Ticonal' 550, which has a large $(BH)_{max}$ product. Moving-coil motors are suitable for use in equipment where uniform motion, rapid reaction and high efficiency are required, and where the greater expense is not a serious drawback. A number of practical versions are discussed, with their applications in tape recorders, pocket dictating machines, measuring recorders, etc.

Two electromagnetic vibrators

J. Timmerman

Introduction

When a reciprocating motion has to be driven electrically, using a rotary electric motor with transmission is really a rather roundabout way of solving the problem. In such a case the more obvious thing to do is to look for an electromechanical system that produces a to-and-fro movement directly. Such systems, which are usually called 'vibrators', can be based on various principles. One such approach is to use the forces exerted on a piece of soft iron or a permanent magnet in a magnetic field produced by an a.c. current. The piezoelectric effect can also be used to produce the required alternating forces from an a.c. voltage input. Magnetostriction can also be considered as a means of energizing a to-and-fro movement, and another scheme is to use the forces between the two plates of a capacitor fed by an a.c. voltage (capacitive drive).

All the possibilities mentioned above require an a.c. voltage supply. In many cases, the a.c. voltage can simply be taken from the mains. The choice of operating principle depends on the particular application; the *amplitude* that the reciprocating motion is required to give is very important. So far, electromagnetic drives have given the largest amplitudes; magnetostrictive or capacitive drives give a very small amplitude, which severely limits their practical application, while piezoelectric drives occupy an intermediate position. In most cases the amplitude required of the reciprocating motion makes it necessary to choose an electromagnetic drive.

This was also the choice made in the two cases described in this article: one vibrator for an electric shaver [1] and one for a refrigerator compressor. The article deals with a number of calculations and measurements made in the development of these two vibrators, after a more general look at the electromagnetic forces and the spring/mass systems in the two vibrators.

Permanent magnet or soft iron

Although both of the vibrators described here make use of electromagnetic forces, they differ quite fundamentally in design. *Fig. 1a* shows the principle of the vibrator for the electric shaver. This vibrator has a soft-iron yoke which forms a closed magnetic circuit

Ir J. Timmerman was with Philips Research Laboratories, Eindhoven, until his retirement in 1967.

together with a movable armature. The yoke is provided with a winding. In the figure, the armature is shown in its rest position; when current flows in the winding (whatever the direction), the armature will be drawn to a position above the poles of the yoke. With an a.c. supply, this happens twice in each period; the armature
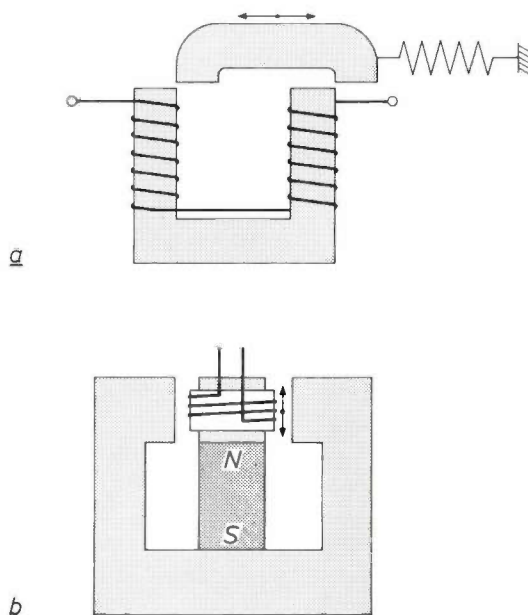
Fig. 1. Two vibrators that make use of electromagnetic forces. *a*) When the coil is actuated, a soft-iron armature is drawn over the poles of the yoke. The force is independent of the direction of the current, so the frequency of vibration is twice the frequency of an a.c. current through the coil. *b*) When current flows in the coil, it moves in the air-gap field of the permanent magnet N, S under the influence of a Lorentz force. When an a.c. current flows in the coil, it vibrates at the same frequency as the current.

will thus vibrate at a frequency equal to twice that of the supply current. It is common practice to make the mechanical resonance frequency of the armature, determined by its mass and the stiffness of the suspension springs, approximately equal to twice the mains frequency so that the amplitude of the vibration will be at a maximum.

The principle of the vibrator for the refrigerator compressor is shown in the diagram of fig. 1*b*. Here a

[1] The idea of an electric shaver with piezoelectric drive was put forward several years ago (see British Patent Application No. 38111/67, registered in 1967 by Mullard Ltd.), but the piezoelectric materials at present available do not have the characteristics required for a successful device.

coil carrying a current moves in the field of a permanent magnet (the magnetic circuit is therefore provided with an air gap in which the coil can move). When current flows in the coil, it is subjected to Lorentz forces, whose direction changes with the direction of the current. With an a.c. supply, the force therefore has the same periodicity as the current; the frequency of the vibration will then be equal to that of the current, and not twice as great (as in fig. 1a). The mechanism of fig. 1b is also used for driving the cone of a loudspeaker. In the refrigerator compressor, the mass of the moving system and the stiffness of the springs are arranged to resonate at the mains frequency.

Electromagnetic forces

The description of the electromagnetic forces can be based on the energy balance of the vibrator: the electrical energy supplied in a time Δt is equal to the sum of the energy dissipated in the windings in that time, the increase in magnetic field energy and the mechanical work due to the simultaneous displacement of the armature or the coil by a distance Δx. This balance leads to an equation in which both electrical and mechanical quantities occur. We can derive the following expression for the total electromagnetic force F_e from this equation, by reasoning similar to that given for the synchronous motor elsewhere in this issue [2]:

$$F_e = i \frac{d\Phi_0(x)}{dx} - \frac{dW_0(x)}{dx} + \frac{\partial}{\partial x} \int_0^i \Phi_i(x,i')di'. \quad (1)$$

In this expression i is the current in the winding, $\Phi_0(x)$ the flux from the permanent magnet enclosed by the winding at zero current, multiplied by the number of turns, $W_0(x)$ the energy of the field produced by the permanent magnet and $\Phi_i(x,i)$ the flux produced by the coil itself. For the integration of $\Phi_i(x,i)$, the value of x is assumed to be constant. This expression is only valid if hysteresis and eddy-current losses are neglected.

A moving-armature vibrator (fig. 1a) has no permanent magnet. The first two terms of (1) will therefore vanish for this vibrator; the force exerted is entirely due to the variation of $\Phi_i(x,i)$ with x. In the moving-coil vibrator (fig. 1b), on the other hand, there is no change in the shape of the magnetic circuit and hence the field energy W_0 remains constant; the second term therefore vanishes. The two remaining terms refer to the two components of the magnetic flux enclosed by the coil: the flux from the permanet magnet, $\Phi_0(x)$, and the flux produced by the coil itself, $\Phi_i(x,i)$. A qualitative picture of the variation of Φ with x and i is given in fig. 2.
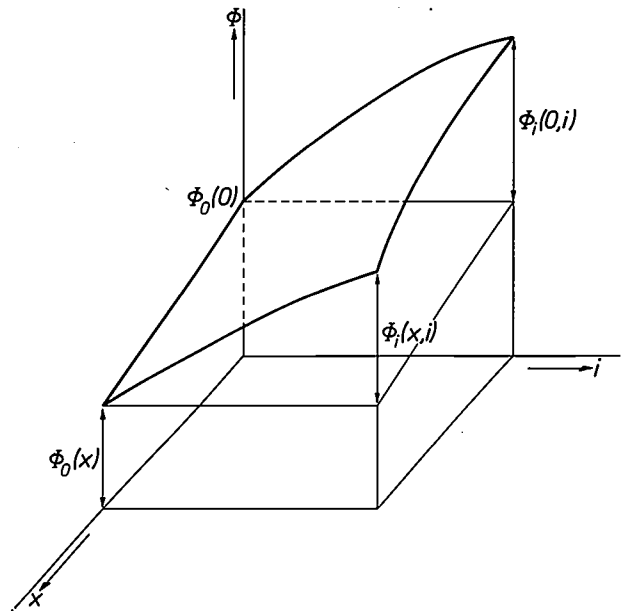
Fig. 2. The magnetic flux Φ linked by the coil as a function of the current i in the coil and the displacement x of the coil (in fig. 1b above). The flux Φ consists of a component Φ_0 from the permanent magnet and a component Φ_i generated by the current through the coil.

The system of masses and springs

In both types of vibrator the moving part is connected to the stator by springs which determine the rest position. When the motor is working, the 'stator' will not be stationary either: the electric shaver is held in the hand, which permits a certain amount of movement, and the compressor is suspended from springs in the refrigerator to minimize the transmission of vibrations. The movement of the stator has an effect on the resonant frequency of the system; this therefore has to be taken into consideration when setting up the vibrator to give the maximum stroke.

If for simplicity we assume that the stator can be represented as connected to a fixed reference point (earth) by means of a spring, then in both cases the entire vibrating system can be represented by the diagram of fig. 3, where m_1 is the mass of the moving part and m_2 the mass of the stator. The masses m_1 and m_2 are connected by a spring of stiffness s_1, and m_2 is connected to earth by a spring of stiffness s_2. The electromagnetic force F_e moves m_1 and m_2 with respect to one another; a frictional force is therefore produced in the element l, the load of the vibrator; it is this frictional force that performs the useful work. The useful displacement in the vibrator is $x_1 - x_2$.

In practice, m_2 is several times larger than m_1. The effect of the (relatively small) movements of the former can be taken into account by replacing the moving mass m_1 by an effective moving mass m which is con-

nected to earth by a single spring of stiffness s_1. The effective mass m is smaller than m_1; as long as s_2 is small, we may write:

$$m \approx \frac{m_1 m_2}{m_1 + m_2}. \tag{2}$$

In practice, therefore, the resonant frequency is slightly higher than that of the system consisting of m_1 and s_1 alone.

To derive equation (2), we start by writing the equations of motion for the masses m_1 and m_2:

$$m_1\ddot{x}_1 + F_l + s_1(x_1 - x_2) = F_e, \tag{3a}$$

$$m_2\ddot{x}_2 - F_l - s_1(x_1 - x_2) + s_2x_2 = -F_e, \tag{3b}$$

where F_l is the frictional force. Adding these equations gives:

$$m_1\ddot{x}_1 + m_2\ddot{x}_2 + s_2x_2 = 0, \tag{4}$$

a result which we shall be using below. It is important for the solution of the equations that the motions are practically sinusoidal, as is found in practice. The loading force F_l does not vary sinusoidally, but it can be replaced by a sinusoidal variable by expanding it in a Fourier series and taking only the first term (the fundamental) of this expansion. Such a procedure is allow-
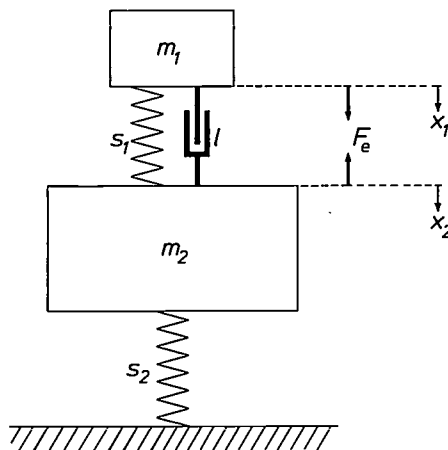
Fig. 3. The mechanical quantities that determine the behaviour of the two vibrators. m_1 mass of the moving part. s_1 stiffness of the springs connecting the moving part to the stator. l frictional load on which the useful work is performed. F_e electromagnetic driving force. m_2 mass of the stator. s_2 stiffness of the connection between the stator and earth; for simplicity, this connection is replaced here by a spring.

able because this term is the only one with the same period as the motion, and therefore the only one that contributes to the work done. The fundamental can be regarded as a force due to fluid friction, which is proportional to the velocity of m_1 relative to m_2. Substitution in (3) gives:

$$m_1\ddot{x}_1 + r(\dot{x}_1 - \dot{x}_2) + s_1(x_1 - x_2) = F_e,$$

where r represents the fluid friction.

Introducing the effective mass m, we find:

$$m(\ddot{x}_1 - \ddot{x}_2) + r(\dot{x}_1 - \dot{x}_2) + s_1(x_1 - x_2) = F_e.$$

This is the equation of motion for a system with a single mass m

and a spring stiffness s_1. The angular resonant frequency ω_r of this system is given by:

$$\omega_r = \sqrt{\frac{s_1}{m}}.$$

This must be equal to the angular frequency at which the actual system consisting of m_1, s_1 and m_2 resonates. The magnitude of m then follows from:

$$m(\ddot{x}_1 - \ddot{x}_2) = m_1\ddot{x}_1$$

and (4). By assuming simple harmonic motion at an angular frequency ω, we find the following expression for the effective mass:

$$m = m_1 \frac{\omega^2 m_2 - s_2}{\omega^2(m_1 + m_2) - s_2}.$$

This shows that the value of m depends on ω. In practice, the spring that connects the system to the earth is generally quite weak, i.e. s_2 is small. If s_2 is neglected in the above equation we get the expression for m given in equation (2).

Vibrator for an electric shaver

The moving armature of the vibrator of fig. 1a is made of laminated soft iron. It moves parallel to the pole faces of the soft-iron yoke, a short distance (e.g. 0.2 mm) away from them. Depending on the construction, the movement is either linear or along the arc of a circle; in the second case the armature oscillates about a pivot. This arrangement is often used in practice (*fig. 4*), but we shall restrict ourselves here to the linear motion, since the path of the motion makes no essential difference to the calculation of the forces and amplitudes.

The friction to be overcome by the vibrator in a shaver or clipper has the nature of dry or Coulomb friction. This means that the magnitude of the frictional force is independent of the velocity, and that the frictional force changes direction with the motion.

The magnitude of this frictional force is an important parameter in the design of a vibrator. One of the first questions that arise is how the stroke and the mechanical power supplied vary when the frictional force changes.

The sensitivity of the armature amplitude to changes in the mechanical adjustment or, conversely, in the frequency of the a.c. voltage is also of importance. The answer to these questions has been sought by means of measurements on test models and by solving the differential equations that describe the behaviour of the vibrator. No analytical solution can be found for these equations, because of the nonlinear character of the frictional force; they can however be solved numerically, and simulation on an analog computer is also possible. Both these approaches have been tried.

[2] E. M. H. Kamerbeek, Electric motors, this issue, p. 215, in particular p. 222.
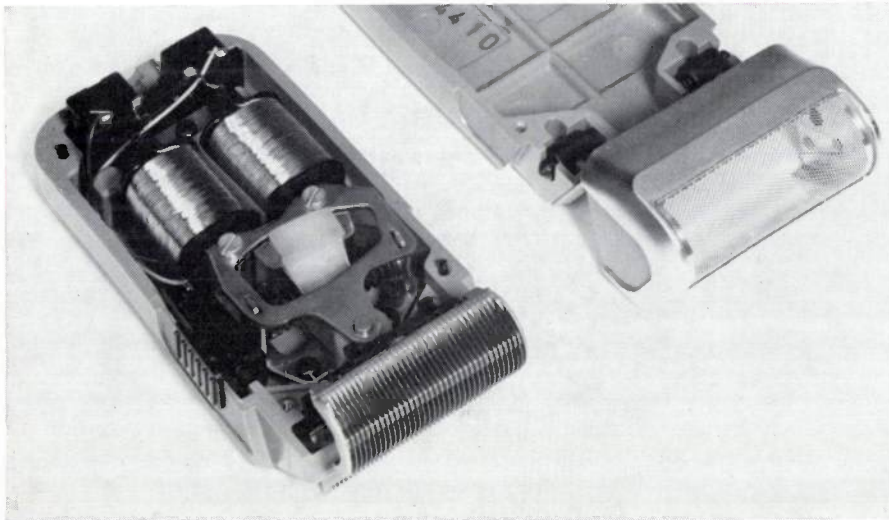
Fig. 4. Vibrator-driven electric shaver. The armature, which moves along an arc of a circle, can be seen below the coils. The armature moves a plastic lever that carries the shaver cutters at its other end. The stiff springs that tune the armature to a frequency of about 100 Hz can be seen just beneath it.

The differential equations

One of the differential equations is the dynamic equation of motion. This was given above in terms of the effective mass m (also defined above); the displacement of the armature with respect to the stator is denoted by x (fig. 5). If we denote the constant force of the Coulomb friction by F_{fr}, the equation of motion may be written:

$$m\ddot{x} + F_{fr} \frac{\dot{x}}{|\dot{x}|} + s(x - x_0) = F_e, \qquad (5)$$

where x_0 is the displacement of the armature in the rest position.

Since there is no permanent magnet in the present case the electromagnetic force F_e is given by the last term of equation (1):

$$F_e = \frac{\partial}{\partial x} \int_0^i \Phi_i(x,i')di'. \qquad (6)$$

The current i appears in this equation; in practice, this is not given so the a.c. voltage v across the terminals has to be used. The current i can be obtained from this with the aid of the voltage equation

$$v = iR + \dot{\Phi}_i, \qquad (7)$$

where R is the resistance of the winding.

To complete this set of equations, we need an expression for Φ_i as a function of x and i. We have determined this relation experimentally by measurements on a test model, which had two identical coils arranged as a bifilar winding on the same former. Reversal of a known d.c. current I in one part of the winding leads to an induced voltage pulse in the other; Φ_i can be determined from this pulse. If this procedure is repeated at various positions of the armature, Φ_i is obtained as a function of x and i. The measured values

are plotted in fig. 6 as a family of Φ-x curves with I as parameter. It can be seen that when the armature deviation exceeds the pole width b, there is very little further variation of the flux with the position of the armature; $\partial\Phi_i/\partial x$ is thus very small in this range, and so is the force.
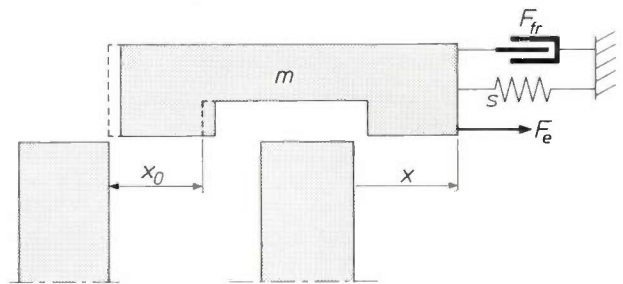
Fig. 5. The quantities in the equation of motion of the armature. x displacement. x_0 displacement in the quiescent position. m effective moving mass. s stiffness of spring. F_{fr} frictional force. F_e electromagnetic driving force.
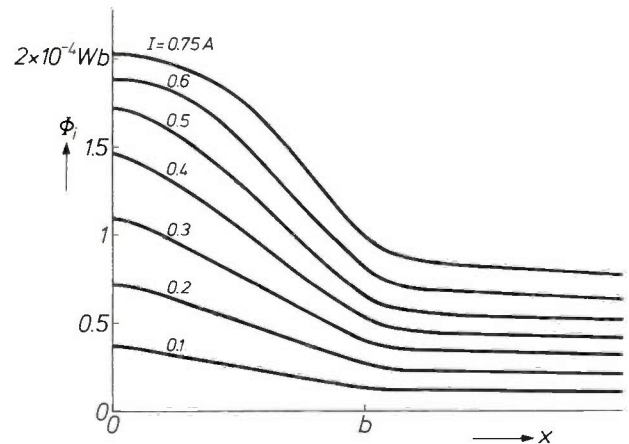
Fig. 6. The magnetic flux Φ_i linked with the winding as a function of the displacement x of the armature. The graph is based on measurements carried out on a test model by reversing the direction of a known d.c. current I and measuring the voltage pulse induced by this; I is a parameter in the figure. When the armature displacement exceeds the pole width b, the flux through the winding has a low, constant value.

The curves of fig. 6 are represented to a good approximation by the equation

$$\Phi_i = A \left[e^{-\alpha x^2} + e^{-\alpha x^4} \right] + B, \qquad (8)$$

where A and B are functions of i, and α is a constant. This approach thus gives a set of equations which completely describe the vibrator.

The function B is simply a constant times i; this constant therefore has the dimensions of inductance and represents the inductance of the winding when the yoke is not closed by the armature (see the flat portion of the curve for $x > b$ in fig. 6). For A we can write: $A = a_1 i - a_2 i^3$; the coefficient a_1 is equal to the inductance added when the armature arrives above the poles of the yoke, while the term $-a_2 i^3$ represents the magnetic saturation.

Amplitude of vibration, delivered power

A numerically computed solution of equations (5-8) is given in *fig. 7*. This solution is based on a number of assumptions: that the supply voltage is sinusoidal, the mechanical system is tuned to twice the frequency of this supply voltage, and the frictional force is such that the system delivers the maximum mechanical power. One striking feature of fig. 7 is that the motion is practically sinusoidal, even though the force is far from sinusoidal. This indicates that the mechanical system is very lightly damped at the optimum friction, and behaves as a resonator with a high Q (quality factor).
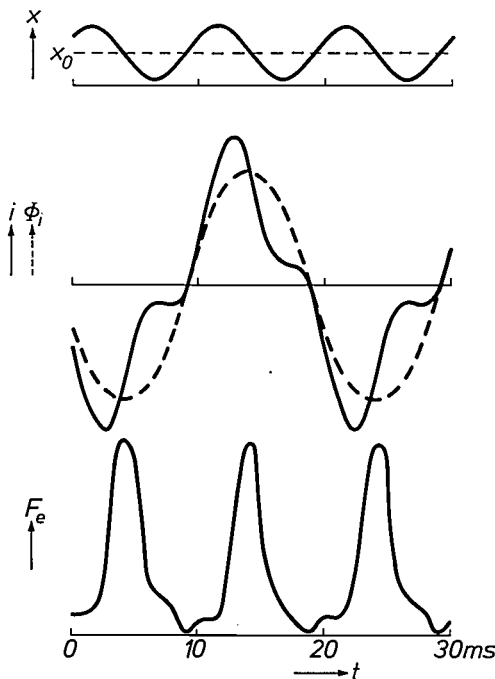
Fig. 7. Variation of the armature displacement x, the current i, the magnetic flux Φ_i and the electromagnetic force F_e as functions of the time t. The curves are the result of a numerical calculation. It can be seen that the motion of the armature is practically sinusoidal.

A system that only has Coulomb friction as described by (5) and driven by a force F_e of sinusoidal waveform will have an infinite amplitude at resonance. In fact this amplitude is not infinite, since the waveform of the force deviates considerably from a sinusoid because of the geometry of the vibrator. As we shall see below, the force can even be negative during part of the cycle.

The amplitude x of the motion of the armature is highly dependent on the choice of the rest position x_0.
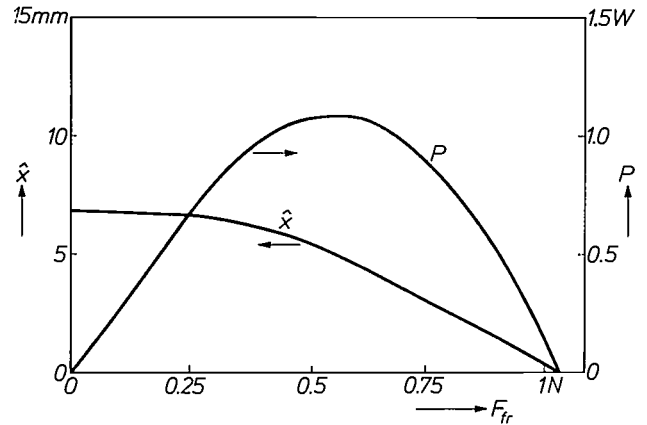
Fig. 8. The amplitude $\hat{x}$ of the armature movement, and the delivered mechanical power P, as functions of the frictional force F_{tr} experienced by the armature. The graph applies to a model with a pole width of 5 mm; the displacement in the quiescent position has been made equal to the pole width. Even without a frictional load, the amplitude never becomes much greater than the pole width. The delivered power P is calculated for a vibration frequency of 100 Hz; it has a definite maximum as a function of F_{tr}.

It has been found that with the type of vibrator considered here, the amplitude is at a maximum when x_0 is approximately equal to the pole width (see fig. 5).

We have also calculated the amplitude of the armature as a function of the frictional force for a vibrator with a pole width of 5 mm (see *fig. 8*). Even when the frictional force is zero, the amplitude is found to be little greater than the pole width. Simulated curves of the force (*fig. 9*) obtained with the analog computer show that at low frictional loads the force becomes negative during part of the cycle; this apparently leads to the amplitude limitation mentioned above.

Since the frictional force is constant, the mechanical power delivered can easily be calculated. During one cycle of the motion of the armature, the distance travelled is $4\hat{x}$, so the work done is $4\hat{x}F_{tr}$. This work is performed $2f_m$ times per second, where f_m is the mains frequency, so the mechanical power P may be written:

$$P = 8 f_m \hat{x} F_{tr}.$$

This power is also plotted in fig. 8 for a 50 Hz supply. It shows a marked maximum of about 1 W at an armature amplitude slightly less than the maximum amplitude, and approximately equal to the pole width. At
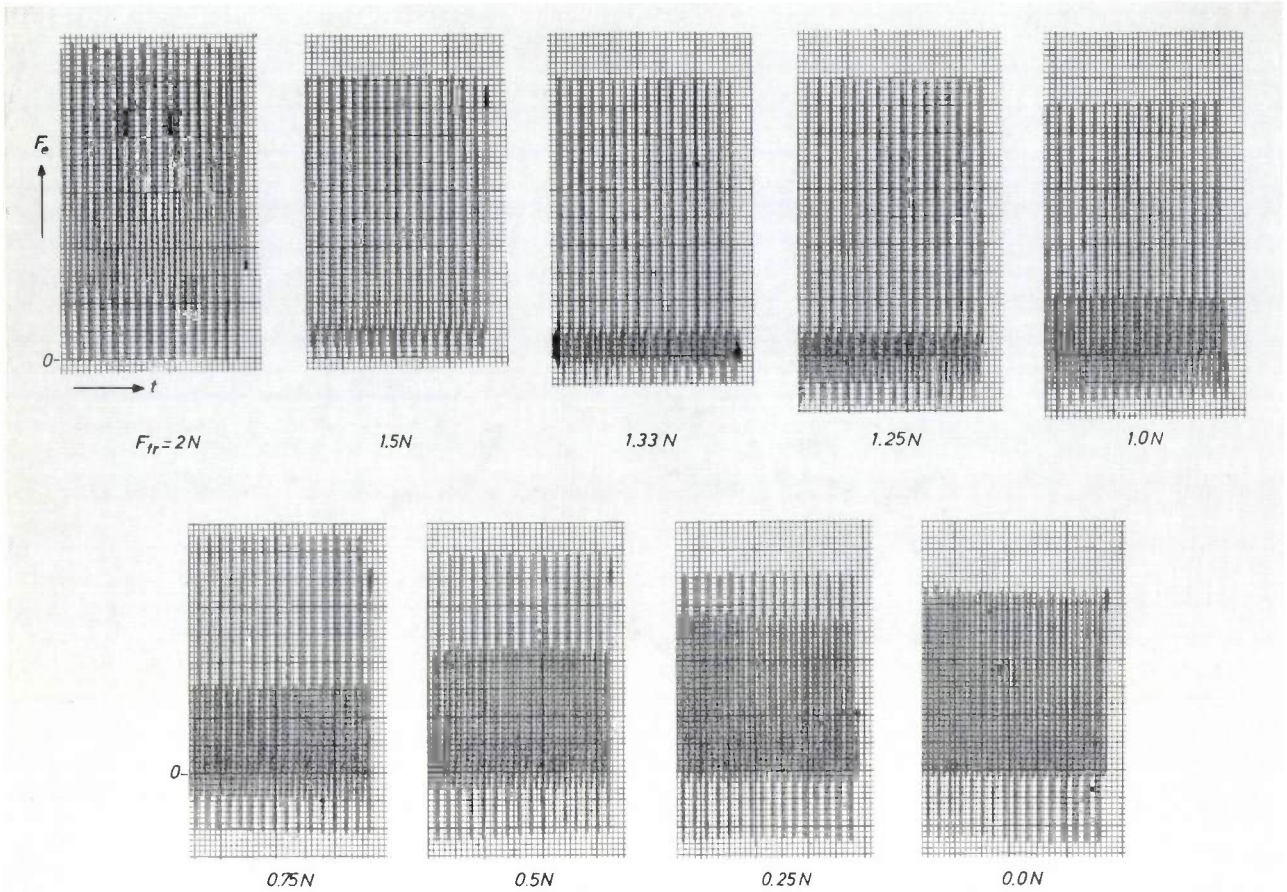
Fig. 9. The variation of the electromagnetic force F_e for various frictional loads F_{fr}. At small values of F_{fr}, F_e is negative during part of each period; this limits the amplitude of the motion, even when the load is small or zero. The curves were simulated on an analog computer.

this amplitude the variation in flux during the stroke is at its maximum, since the armature just covers the poles completely in one of the extreme positions.

Voltage variations

Variations in the supply voltage from the nominal value while the shaver is in use (and variations of ± 10% in the mains voltage can indeed be expected) must not affect the operation of the shaver too greatly: too high a voltage should not make the armature exceed the limits of its free path, and too low a voltage should not make the shaving speed too low. The amplitude and power have therefore also been calculated for voltages 10% above and below the nominal value. Since these calculations were carried out for a test model with stepped poles (fig. 10), which gives a higher drive force and a lower amplitude, the numerical results are not comparable with the results of fig. 8. It can be seen (fig. 11) that the maximum amplitude (without frictional load) does not change when the voltage is varied; the amplitude at which the vibrator delivers its maximum power is also found to

be constant. Both these amplitudes are closely related to the dimensions of the vibrator. At the nominal frictional load of 0.82 N, the amplitude only increases slightly when the voltage is increased, but it falls off quite sharply when the voltage is reduced. To counteract this, it is advisable to load the vibrator with slightly less than the nominal frictional force; the amplitude variation resulting from a given voltage variation is then less.

Behaviour away from resonance

It has already been mentioned that the vibrator acts like a resonator with relatively weak damping when the frictional load has the optimum value. It is conse-
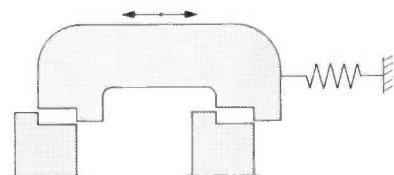
Fig. 10. Stepped poles. A vibrator provided with such poles has about half the amplitude of one with flat poles, but at twice the loading force.
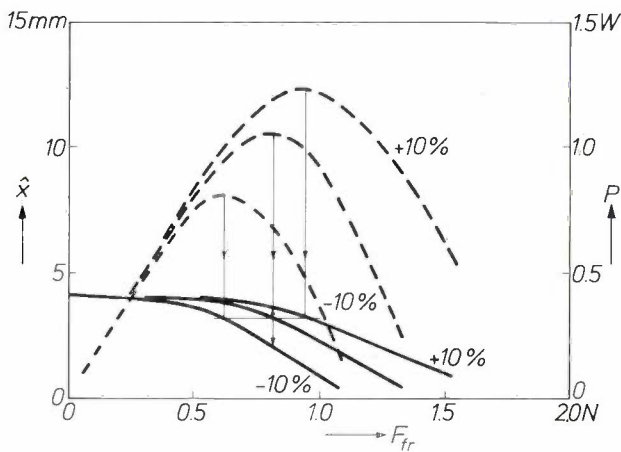
Fig. 11. Armature amplitude $\hat{x}$ and delivered power P at the nominal supply voltage and at voltages 10% above and below this. At the optimum setting of the frictional force F_{fr} (0.82 N), a drop of 10% in the voltage leads to an appreciable drop in amplitude. If the frictional force is adjusted to a lower value, the amplitude drop is reduced. It can be seen that the amplitude at which the maximum power is delivered is the same in each case (3.2 mm).
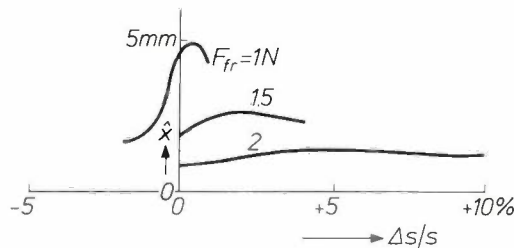
Fig. 12. Effect on the armature amplitude $\hat{x}$ of a variation Δs of a few per cent in the spring stiffness. At $\Delta s = 0$, the armature is tuned to twice the mains frequency. The frictional force F_{fr} is taken as a parameter; the maximum power is delivered at a frictional force of 1 N.
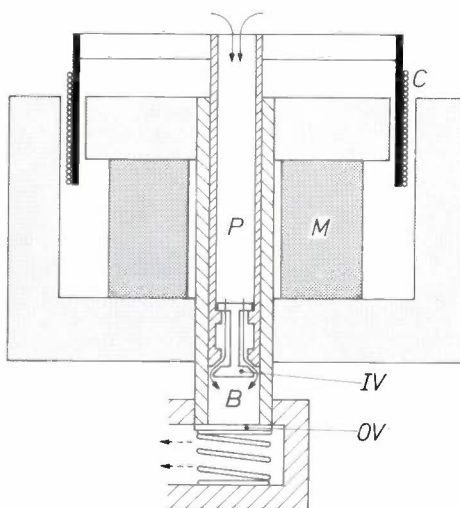
Fig. 13. Cross-section of a vibrator compressor. *C* coil. *P* hollow plunger. *B* cylinder bore. *IV* inlet valve. *OV* outlet valve. *M* magnet.

quently to be expected that the amplitude of the armature motion will exhibit a pronounced maximum at the resonant frequency. This was confirmed by a computer simulation in which the parameter introduced for the spring stiffness s was varied by several per cent (*fig. 12*). The optimum frictional load was 1 N here; at higher damping, the curve is much flatter.

Vibrator for a refrigerator compressor

In a refrigerator, a coolant (usually 'Freon') is pumped round the cooling circuit by a compressor which takes the coolant in at its inlet as a gas, and compresses it. The compressor is situated in a tank filled with the coolant, completely isolated from the outside air atmosphere. The pressure in the tank is equal to the intake pressure of the pump. The coolant is compressed by means of a plunger that moves to and fro in a cylinder. Refrigerator compressors are normally driven by induction motors, with the rotary motion converted into a reciprocating motion for the plunger.

It seemed likely that use of a vibrator to drive the plunger might lead to a simpler and more efficient construction. To test this assumption in practice, we have developed a compressor for a small domestic refrigerator, driven by a moving-coil vibrator as illustrated in fig. 1*b*. The motor and compressor were designed as a single unit: the plunger and its cylinder are situated centrally in the hollow magnet (*fig. 13*).

The force exerted by the gas had to be taken into account in the design of this compressor. For example, in determining the stiffness of the springs required to bring the moving masses of coil and plunger to resonance at the mains frequency, the force exerted by the compressed gas in the cylinder on the plunger is of significance. This gas volume behaves as a nonlinear spring, since the force exerted is not proportional to the displacement of the plunger.

Our vibrator compressor, shown in section in *fig. 14*, is driven by two 'balanced' coils. These coils move in two air gaps. Since the field of the permanent magnet has opposite directions in the two air gaps, the current in the two coils must also flow in opposite directions. The fields developed by the two coils therefore largely cancel each other, ensuring that the permanent magnet is not demagnetized by the current pulse produced on switching on (the amplitude of this pulse can be several times the operating current). Moreover, with this design the coils do not produce an alternating field in the magnetic circuit, so that the iron losses are low.

Another noteworthy feature of fig. 14 is the use of spiral springs. These have the advantage that the
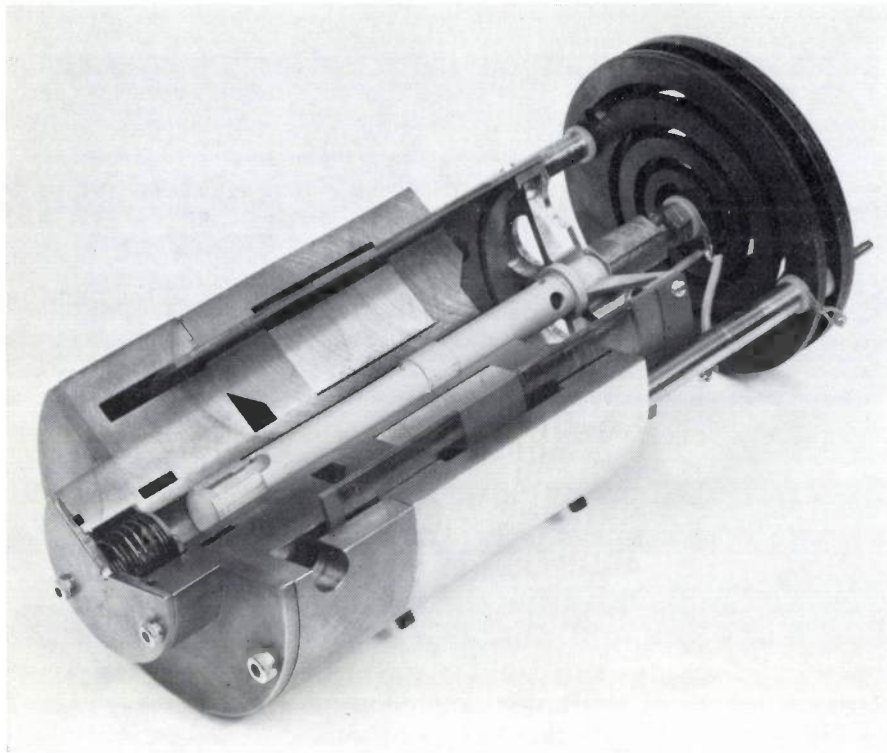
force of the spring acts along the centre-line of the plunger and not at one side, as is the case with helical springs. In the latter case, the friction of the plunger in the cylinder would be too high. Spiral springs can be used in single or double versions (*fig. 15*). They are carefully centred with the aid of an accurately fitting central hole for the plunger mounting and two holes with a large clearance at the edge for attaching the spring to the vibrator housing. With this arrangement the springs do not exert any lateral force on the plunger.

It has been found that satisfactory vibrator compressors can be built on the basis of the design illustrated in fig. 14. We shall now discuss the electromagnetic and compression forces in these compressors in somewhat greater detail.

Electromagnetic forces

In the discussion of the general expression (1) for the electromagnetic forces, it has already been mentioned that in the moving-coil vibrator the energy of the permanent-magnet field is not affected by the motion of the coil, so that the second term on the right-hand side is zero and the expression for the electromagnetic force is given by:

$$F_{\mathrm{e}} = i \frac{\mathrm{d}\Phi_0(x)}{\mathrm{d}x} + \frac{\partial}{\partial x} \int_0^i \Phi_i(x,i')\mathrm{d}i', \qquad (9)$$

where $\Phi_0(x)$ is the flux from the permanent magnet
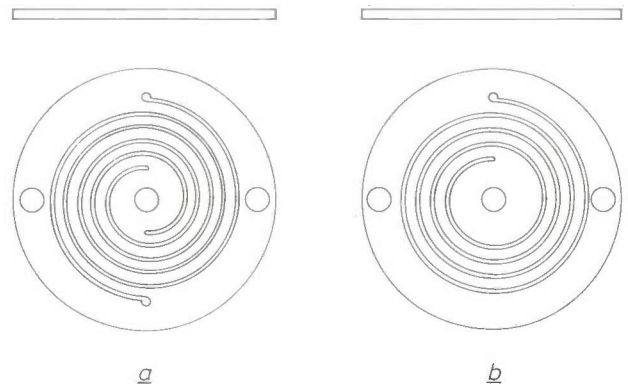
Fig. 15. Double (*a*) and single (*b*) spiral springs.

linked with the coils. By making the coil a little longer (0.5 to 0.7 times the stroke of the plunger) than the height of the pole shoe, $\Phi_0(x)$ can be made to vary almost linearly with the displacement, so that $\mathrm{d}\Phi_0(x)/\mathrm{d}x$ is constant. The force represented by the first term has the same period as the current i, i.e. in practice the same period as the mains voltage.

The second term can be simplified, since measurements have shown that the flux Φ_i is to a good approximation proportional to the current i; this is due to the fairly large air gaps, which restrict the effects of saturation of the magnetic circuit. The second term may therefore be written $\frac{1}{2}i^2\mathrm{d}L(x)/\mathrm{d}x$. This produces second harmonics in the force, but measurements have shown that the amplitude of this term is 50 to 60 times less than the first term. This means that only the first

term is of importance for the electrical design. The design calculations are based on the work performed in compressing the gas during each stroke of the plunger (the compression work).

Compression work

When the plunger is moving downwards, the gas beneath it is compressed until the gas pressure is high enough to open the outlet valve against the force of the valve spring plus the gas pressure on the other side of the valve (see fig. 14). During the rest of the stroke the pressure stays at this value (the compression pressure), and gas is pushed out of the cylinder. When the plunger reaches its lower dead point, there is still a small amount of gas left in the unswept volume. When the plunger starts to move upwards, the outlet valve closes and this residual gas expands. Its pressure soon falls below the intake pressure, so that the inlet valve at the bottom of the plunger opens and the cylinder is connected to the 'Freon' in the tank via the hollow plunger. The rest of the upward movement of the plunger has no effect on the gas pressure. When the upper dead point is passed, the inlet valve (which is freely movable in the plunger) closes under its own inertia, and a new compression stroke starts.

If the vibrator is driven at 50 Hz, the whole compression and expansion cycle is completed in 20 ms. This is so fast that we can assume that both compression and expansion are adiabatic. We can therefore write:

$$p(x) = p_0 \left(\frac{V_0}{V(x)} \right)^{\varkappa}, \qquad (10)$$

where p_0 is the intake pressure and V_0 the gas volume under the plunger at the moment when the inlet valve closes; $p(x)$ and $V(x)$ are the pressure and the volume for a displacement x of the plunger, and $\varkappa$ is the ratio of the specific heats of the gas at constant pressure and constant volume. If the pressure $p(x)$ is plotted as a function of the volume $V(x)$ in a p-V diagram the result is a closed contour, which is traversed once each time the plunger completes a stroke (*fig. 16*). The pressure p_1 is the compression pressure. The straight sides of the contour shown in fig. 16 are approximations; the actual curve will be more complicated, because of vibration of the outlet valve and the inertia of the inlet valve. However, the straight lines make it easier to calculate the volume enclosed by the curve, which represents the compression work performed by the plunger per stroke:

$$W = - \oint p \, \mathrm{d}V.$$

Making use of (10) for the curved parts of the contour,

we obtain the following expression for W:

$$W = p_0 V_0 \frac{\varkappa}{\varkappa - 1} \left\{ 1 - \frac{V_2}{V_0} \left(\frac{p_1}{p_0} \right)^{\frac{1}{\varkappa}} \right\} \left\{ \left(\frac{p_1}{p_0} \right)^{\frac{\varkappa - 1}{\varkappa}} - 1 \right\}, \quad (11)$$

where V_2 is the unswept volume. This compression work W is plotted as a function of the ratio V_2/V_0 in *fig. 17*, with the ratio p_1/p_0 of compression pressure to intake pressure as parameter. This family of curves applies to 'Freon' 12 (difluorodichloromethane, CF_2Cl_2), the coolant gas normally used in refrigerators, for which $\varkappa = 1.13$.
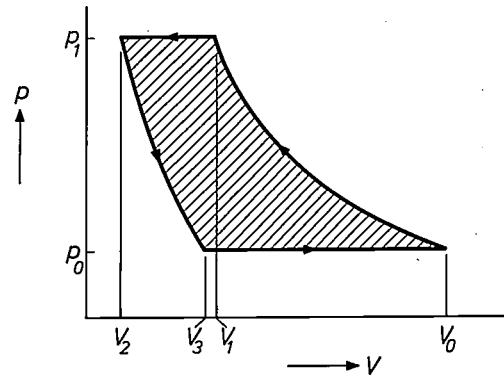
Fig. 16. The compression work performed during each stroke of the plunger is equal to the area enclosed by the contour in the p-V diagram; this contour is for adiabatic compression. V_0 volume under the plunger at the start of compression; when this volume has been compressed to V_1, the outlet valve opens. The gas is now expelled, apart from the unswept volume V_2. When this volume has expanded to V_3, the inlet valve opens. p_0 intake pressure. p_1 compression pressure.
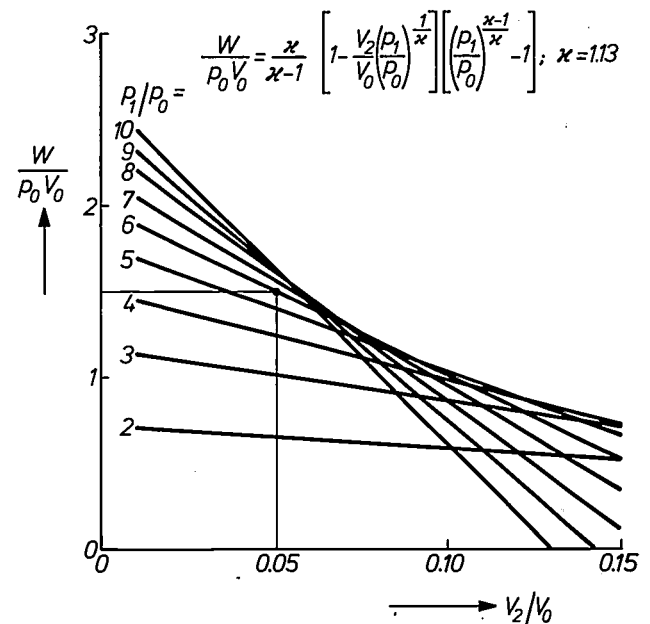
Fig. 17. The compression work W performed during each stroke of the plunger as a function of the expansion ratio V_2/V_0; W is normalized by dividing it by $p_0 V_0$. The shape of the curve varies with the value of the ratio of compression pressure p_1 to intake pressure p_0. The operating point of one of the experimental models is indicated in the figure.

The operating point of one of our test models is situated on this graph at $V_2/V_0 = 0.05$ on the curve $p_1/p_0 = 6$; it can be seen from the graph that the work per compression stroke is then $W = 1.5\,p_0V_0$. Assuming an intake pressure of $p_0 = 2$ bars and an initial volume $V_0 = 3.5$ cm^3, we arrive at a compression power of 52.5 watts at a drive frequency of 50 Hz. These figures all apply to the refrigerating power required for a domestic refrigerator with a capacity of about 120 litres. If the frictional losses are estimated at about 15 to 20 watts, the total mechanical power required is found to be about 70 watts; this has to be supplied by the vibrator.

Power of the vibrator

The design calculations for the vibrator are greatly simplified by the experimental observation that the movement of the plunger is practically sinusoidal, despite the nonlinearity of the *forces* exerted on the plunger by the gas during compression. As in the case of the shaver vibrator, this must come about because of the high Q of the resonating spring/mass system, here formed by the mass of the coil and the plunger and the restoring force of the spiral springs and the enclosed volume of gas. If we restrict ourselves to the first term in expression (9), the same is true of the force, since $d\Phi/dx$ is constant in a well designed vibrator. The mechanical power delivered by the motor is given by the product of force and velocity; both these variables vary sinusoidally with time and are in phase with one another, since the system is resonant. The amplitude of the velocity is $\omega_\mathrm{m}\hat{x}$, so the internal power is

$$P = \tfrac{1}{2}\hat{F}_\mathrm{e}\omega_\mathrm{m}\hat{x} = \tfrac{1}{2}\frac{d\Phi_0}{dx}\hat{\imath}\omega_\mathrm{m}\hat{x}. \qquad (12)$$

The quantity $d\Phi_0/dx$ can be expressed in terms of the magnetic flux density and the length of the wire wound in the coil. If we consider a coil with a single turn of length πd (see *fig. 18*), moving in a radial air-gap field of flux density B, then when this coil moves downward a distance Δx the magnetic flux it encloses will increase by $B\pi d\Delta x$. If the coil has N turns, the flux linked will be N times as great. If the coil is situated partly outside the air gap, as in our vibrator, we can approximate to the result by considering only the windings in the air gap, and multiplying their number by a correction factor γ, which is found in practice to have values between 1.05 and 1.1. It follows that $d\Phi_0/dx = \gamma lB$, where l is the length of wire in the air gap.
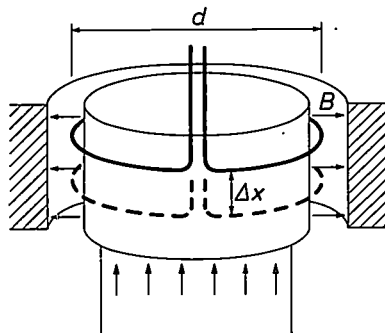
The power is thus given by

$$P = \tfrac{1}{2}\,\gamma lB\hat{\imath}\omega_\mathrm{m}\hat{x}.$$

This expression forms the basis for the electrical design calculations.

For the purposes of these calculations, P and $\hat{x}$ are given by the compressor characteristics; the mains voltage is also given. The design work consists in finding the best compromise between efficiency (minimum losses in the coil), cost price (a higher flux density B requires a more expensive magnet) and reliability (the heat dissipated in the coil must not raise its temperature above a certain limit). A good compromise, which we have used in a test model, is found with a wire length (in the air gap) of 152 m, an air-gap flux density of 0.33 teslas and a current amplitude of 0.57 A (rms value 0.4 A). These values are for the compressor mentioned above for a 120 litre domestic refrigerator; the plunger amplitude is here 14 mm, and with a delivered motor power of 70 W the power required is 88 W. Taking the compressor power as 52.5 W (see above), the overall efficiency of the vibrator compressor is 60%; this compares favourably with the values of 50 to 55% found with conventional refrigerator compressors.

Parasitic oscillations in the springs

The individual turns of the spiral springs have their own resonant frequency, which depends on their mass and stiffness. If one of these resonant frequencies lies near a harmonic of the mains frequency, the turn in question will be brought into resonance by one of the harmonics of the plunger motion (even if this is very small). This effect is sometimes so severe in practice that the spring breaks.

Parasitic oscillations can be suppressed by means of a dynamic absorber, usually consisting of a flat spring mounted on the top of the plunger and carrying little weights at its ends (*fig. 19*). These weights are chosen so that the spring/mass system of the dynamic absorber is tuned to the frequency to be suppressed. At this frequency, the dynamic absorber presents such a high mechanical impedance to the plunger that plunger movement at this frequency is practically eliminated; as a result, the parasitic resonance of the spring is also suppressed.

This effect can be demonstrated theoretically by an analysis in which damping is neglected. Let the moving mass of the vibrator be m and the stiffness of the spiral springs s, and similarly the

Fig. 18. If a single-turn coil moves downwards by a distance Δx, the magnetic flux enclosed by the coil increases by $B\pi d\Delta x$.

moving mass of the dynamic absorber m_d and its spring stiffness s_d; the entire system can then be schematically represented as shown in *fig. 20*, and described by the two differential equations:

$$m_d \ddot{x}_d + s_d(x_d - x) = 0,$$
$$m\ddot{x} + sx - s_d(x_d - x) = F_e. \tag{13}$$

For sinusoidal motion at an angular frequency ω, these differential equations give the two algebraic equations:

$$-\omega^2 m_d x_d + s_d(x_d - x) = 0,$$
$$-\omega^2 mx + sx - s_d(x_d - x) = F_e,$$

from which the following expression for the displacement of the plunger can be derived:

$$x = \frac{F_e(s_d - \omega^2 m_d)}{(s + s_d - \omega^2 m)(s_d - \omega^2 m_d) - s_d{}^2}. \tag{14}$$

At the angular frequency $\omega_d = \sqrt{s_d/m_d}$, to which the dynamic
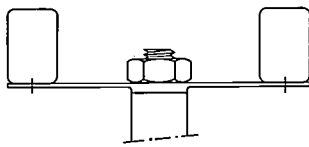
Fig. 19. A dynamic absorber. Two weights at the end of a flat spring mounted on the top of the plunger suppress movements of the plunger at the frequency to which they are tuned.
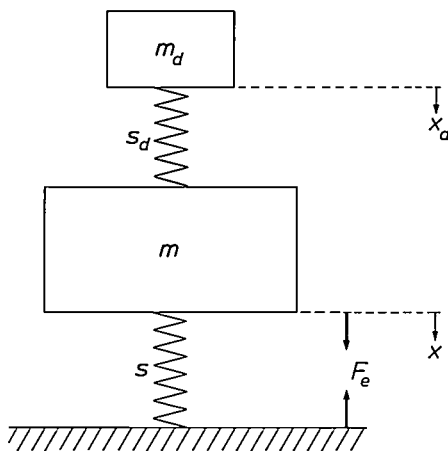
Fig. 20. Mechanical quantities in a vibrator compressor with dynamic absorber. m mass of coil and plunger. s stiffness of the spiral springs. m_d mass and s_d stiffness of the dynamic absorber. F_e driving force.

absorber is tuned, the numerator of this fraction is zero; it follows that plunger motion at the frequency $\omega_d/2\pi$ is impossible, even though the force F_e may contain components of this frequency.

The dynamic absorber inevitably includes a certain amount of damping. This gives rise to a damping term that remains in the numerator of (14) when $s_d - \omega^2 m_d$ vanishes. To ensure efficient operation of the dynamic absorber, its damping must be made as low as possible.

The vibrator has proved to be a suitable power source for driving a refrigerator compressor; its efficiency is better than that of a compressor with an induction motor. The double-coil construction offers an adequate solution to the problems otherwise caused by demagnetization and iron losses, especially at higher powers. Whether vibrator compressors can be used on a commercial scale in refrigerators depends largely on the price for which they can be made. Large magnets are expensive, and can make the vibrator compressor for high powers more expensive than the conventional compressor with rotary drive. At low powers, on the other hand (e.g. 20 W, which corresponds to a 25 litre refrigerator), the simpler construction of the vibrator compressor makes it the best choice.

Summary. Vibrators to provide a direct electrical drive for reciprocating motion can be based on various principles; in practice, however, they usually make use of electromagnetic forces. Two electromagnetic vibrators are described, one with a moving soft-iron armature that is attracted to the poles of a yoke when a coil is energized, and the other with a coil moving in a constant magnetic field. In the first case the vibrator oscillates at twice the drive frequency; this model is intended for use in an electric shaver. The moving mass is made to resonate at twice the mains frequency by means of stiff springs; the motor delivers a mechanical power of more than 1 W. The other vibrator has been designed to drive a refrigerator compressor. In this type, two coils in which the current flows in opposite directions move in the field of a permanent magnet; the fields produced by the coils cancel out, so that demagnetization and iron losses are eliminated. The coils are rigidly connected to a plunger; spiral springs ensure a centrally acting spring force, and bring the whole system into resonance at the mains frequency. The motion of both vibrators is approximately sinusoidal, despite the non-sinusoidal form of the loading forces.

260

Philips tech. Rev. 33, 260-271, 1973, No. 8/9

Frequency-analog speed control

H. Kalis and J. Lemmrich

General

An electronic method of control that has been known since the early thirties is that of the phase-lock-loop technique. In this method the phase angle of a periodic signal whose frequency corresponds to the value of the quantity to be controlled is continuously compared with the phase angle of an applied signal whose frequency represents the specified value. A change in the phase difference between the two produces an error signal that counteracts the change; in this way the two frequencies are permanently synchronized and the quantity to be controlled is kept to the required value. A control system of this type, in which the quantities are represented in the form of a frequency, is referred to as frequency-analog control at Philips Forschungslaboratorium Hamburg [1]. The principle of frequency control is convenient for the control of many kinds of quantity, but is particularly suitable for controlling the speed of electric motors, where the quantity to be controlled is essentially periodic in nature.

We shall start by way of introduction with such a system; see the diagram of *fig. 1*. A voltage V_{ref} of sinusoidal waveform and angular frequency ω_{ref} (or the sum of a number of signals of different frequencies) is multiplied by a voltage of sinusoidal waveform and angular frequency ω in a multiplier circuit M. The output voltage of the multiplier is applied through a lowpass filter LP to an oscillator VCO whose frequency can be controlled by a voltage; this in turn supplies the voltage V of angular frequency ω.

The multiplication produces a voltage V_0 at the output of the multiplier, given by:

$$V_0 = \tfrac{1}{2} ABC\{\cos(\omega_{ref}t\; \omega t - \phi) - \cos(\omega_{ref}t + \omega t + \phi)\}.$$

Here A and B are the amplitudes of the two voltages, C is the multiplication constant and ϕ is the phase of the voltage V with respect to the reference voltage V_{ref}. The output signal of the multiplier thus contains the sum and difference frequencies of the input signals.

The lowpass filter suppresses all frequencies above the cut-off frequency $\omega_c/2\pi$. It always suppresses the sum frequency, and also the difference frequency if this is above $\omega_c/2\pi$. If however the frequency of the controlled oscillator is set close to the reference frequency

ω_{ref}, e.g. by the auxiliary voltage V_1, so that the difference $|\omega_{ref} - \omega|$ is less than ω_c, then synchronization is obtained automatically and ω becomes equal to ω_{ref}. Synchronization is initiated because the control signal V_a increases if the difference frequency becomes smaller; the frequency of the controlled oscillator is therefore pulled in to the reference frequency. Once the system is synchronized the multiplier circuit acts as a rectifier, controlled in phase by the oscillator, which takes a phase difference such that

$$\omega_{ref} = k\{\tfrac{1}{2} ABC \cos(-\phi) + V_1\}, \qquad (1)$$

where k is a generator constant.

It can be seen from the diagram of fig. 1 that this method of obtaining synchronization depends on the existence of a control loop. The error signal in this loop is given by the phase difference ϕ between the two input signals while the angular frequency ω is the controlled quantity [2].

Now let us consider a phase-lock-loop system in which the voltage-controlled oscillator VCO is replaced by a motor whose speed of rotation is determined by the voltage V_a. The shaft of the motor is assumed to be connected to an a.c. tachogenerator that gives an output signal at a frequency proportional to the angular velocity ω corresponding to the speed of rotation; this signal corresponds to the voltage V in fig. 1. With such a system there could be no possible error between the speed of the shaft — represented by the angular frequency ω — and the specified value ω_{ref}: the speed control would therefore be ideal. The idea has been put into practice in various places in the last ten years or so [3], but had been used as early as 1938 [4].
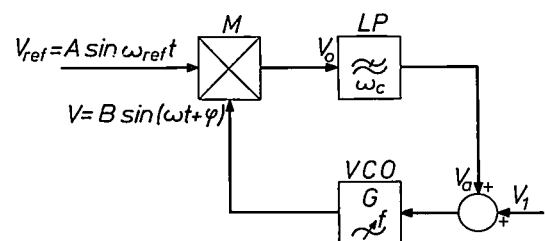
Fig. 1. Phase-lock loop for synchronizing the frequency of a voltage controlled oscillator VCO with a reference frequency $\omega_{ref}/2\pi$. The output voltage V of the oscillator is multiplied by the reference signal V_{ref} in the multiplier circuit M and the resultant signal is passed through a lowpass filter LP to give the voltage V_a. This voltage V_a controls the oscillator. An auxiliary voltage V_1 can be applied externally to initiate the synchronization.

Dipl.-Ing. H. Kalis and Dipl.-Ing. J. Lemmrich are with Philips Forschungslaboratorium Hamburg GmbH, Hamburg, West Germany.

Phase-lock loops for speed control of electric motors

In this section we shall look at the system of fig. 1 to see if it is the most suitable one for speed control and to find the best way of replacing the controlled oscillator by a motor fitted with a sensing device. We shall also attempt to illustrate the most important common features of the various versions of the system produced in our Hamburg laboratories by two simple but typical examples.

Since the measured frequencies have to be produced by tachogenerators, the frequencies used · in motor control are relatively low (a few hundred hertz). The signals are usually rectangular pulses, because these are the easiest to produce. Rectangular pulses are used without exception in all the literature that we quote.

Control with the aid of analog signals

In our first example of a frequency-analog speed control the multiplier of fig. 1 is replaced by a phase-detector circuit in which the phase of the pulses is converted into a voltage.

The operation is explained with the aid of *fig. 2*. Fig. 2a shows the reference signal (corresponding to V_{ref}) and fig. 2b the measured signal (corresponding to V) as a function of time; a phase change in the measured signal is produced at time t_0, for example by changing the load on the motor. Fig. 2c shows a sawtooth voltage whose slope angle α is proportional to the reference angular frequency ω_{ref}; the start of the sawtooth always occurs at the leading edge of the reference-signal pulse. The phase difference between the reference and measured pulses is converted into a quasi-steady-state signal $V_a(t)$ by a hold circuit, which holds the instantaneous value of the sawtooth voltage at the instant when a measured pulse appears and produces this voltage at the output as a control signal (fig. 2d). This control signal is independent of the frequency, since the slope angle α of the sawtooth voltage is proportional to the reference frequency.

Fig. 3 shows the complete system, including the motor M, in the form of a block diagram. It is assumed that M is a d.c. motor with separate excitation because

of the extremely convenient static and dynamic equations relating rotor current and motor torque, and also motor torque and speed.

The part of the system corresponding to the voltage-controlled oscillator of fig. 1 — motor, controlled rectifier bridge (CRB) and pulse tachogenerator (L, SD, Pho, PS) — is enclosed by a chain-dotted line. Since there are time delays present inside this block
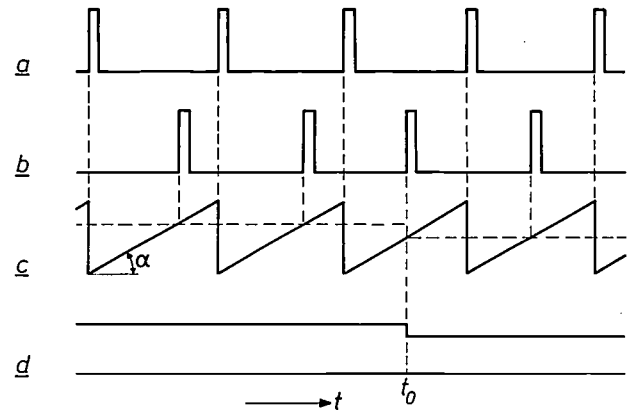
Fig. 2. Conversion of the pulse phase to a voltage. *a*) Reference pulses. *b*) Measured pulses. *c*) Sawtooth voltage whose start always coincides with the leading edge of the reference pulse. When the measured pulse arrives the sawtooth voltage has reached a value proportional to the phase difference between reference and measured pulses. *d*) This value is held in a hold circuit. At time t_0 there is a step in the phase difference, e.g. as a result of a sudden change in the load on the motor.
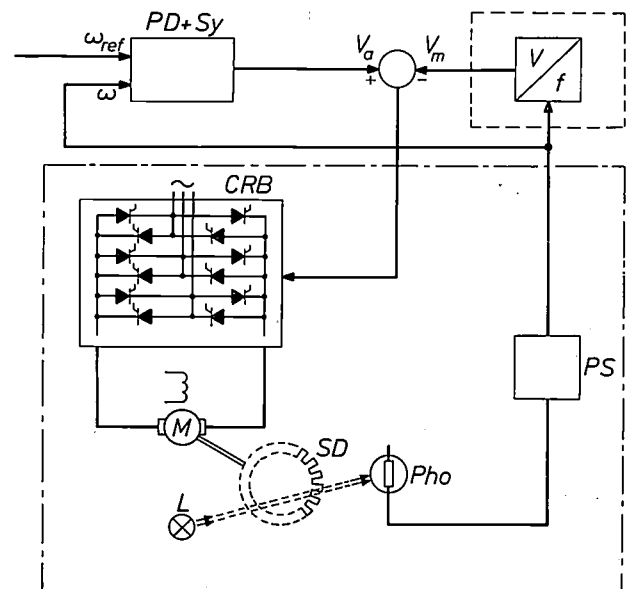
Fig. 3. Frequency-analog control of a d.c. motor M with the speed of rotation measured by a light source L, a slotted disc SD and a photodetector Pho. The rotor current is controlled by a controlled rectifier bridge CRB. This circuit is controlled by the analog output signal of the phase detector and synchronizing circuit $PD + Sy$, after subtraction of the damping signal V_m, which is proportional to the speed and is derived from the photodetector via the pulse shaper PS and the frequency/voltage converter.

[1] D. Gossel, Meßsysteme und Regelungen mit Frequenzsignalen, Messen, Steuern, Regeln 14, 22-28, 1971.
   D. Gossel, Frequenzanalogie — Ein Konzept für Meß- und Regelsysteme mit digitaler Signalverarbeitung, Elektrotechn. Z. A 93, 577-581, 1972.
[2] A general treatment of phase-lock loops is given in: F. M. Gardner, Phaselock techniques, Wiley, New York 1966.
[3] M. J. Campos Costa, Obtaining precise induction motor speed, Control Engng. 10, No. 6, 92-93, 1963.
   J. Lemmrich, Der synchronisierte Induktionsmotor, Elektrotechn. Z. A 85, 724-726, 1964.
   E. K. Howell, Solid-state control for dc motors provides variable speed with synchronous-motor performance, IEEE Trans. IGA-2, 132-136, 1966.
   H. Kalis and J. Lemmrich, Frequenzanaloge Drehzahlregelungen I, II, Regelungstechnik 16, 497-502, 555-562, 1968.
[4] H. Rinia, Television with Nipkow disc and interlaced scanning, Philips tech. Rev. 3, 285-291, 1938.

(the moment of inertia of the rotor and time constant of the rotor circuit) its dynamic behaviour does not correspond to that of the oscillator of fig. 1; a damping signal is required, to prevent instability. This is supplied by a circuit that converts frequency to voltage (the part of the circuit inside the dashed line in fig. 3). More will be said about this circuit later.

The operation of the speed-control system is as follows (leaving aside for the moment the signal from the frequency/voltage converter). The output voltage from the phase detector PD, whose operation has

it also follows that for steady-state conditions

$$T_l = k_1 k_2 \Delta\phi,$$

where k_2 is another gain factor. The phase difference between reference and measured signal is therefore proportional to the load, from which it can be seen that this — and any other frequency-analog-controlled machine — behaves like a synchronous machine [5].

To bring out these relationships more clearly, we shall show how the complete system reacts to a sudden step in the load when it is in equilibrium. After the
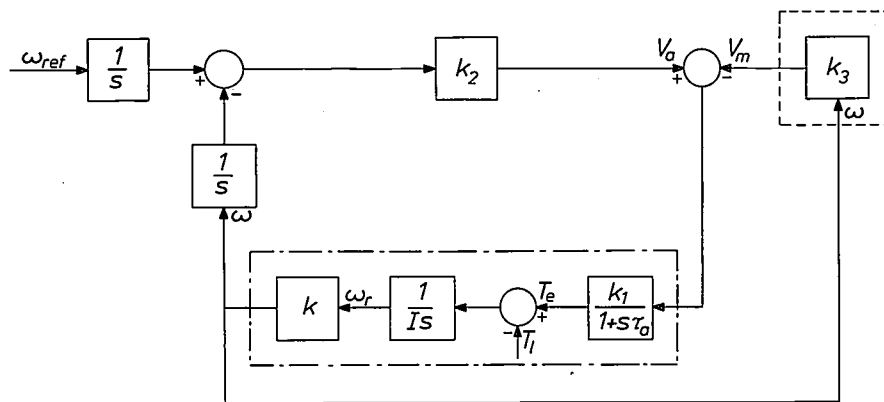
Fig. 4. The block diagram of fig. 3 redrawn in control-theory terms. The transfer function of each block is shown; s is the complex frequency $\sigma + j\omega$.

already been described, functions as a reference value for the rotor-current control circuit of the d.c. motor. This ensures that when the very short-lived transients from the rotor-current control circuit have died away the torque T_e of the motor is directly proportional to V_a. The relation between the angular velocity measured by the sensing device and the torque T_e of the motor is given by the equation:

$$T_e - T_l = I \, d\omega_r/dt = I \, d^2\phi_r/dt^2, \qquad (2)$$

where T_l is the load torque, I the total moment of inertia of the rotating components, ω_r their angular velocity and ϕ_r the instantaneous value of the angular position of the rotor. The sensing device gives a measured frequency

$$\omega/2\pi = k\omega_r/2\pi, \qquad (3)$$

where k is the number of pulses per revolution.

If the phase-lock loop is in equilibrium ($\omega_{ref} = \omega$), then $d\omega_r/dt = 0$ and it follows from (2) that T_e is equal to T_l and therefore, because of the current control mentioned above, equal to $V_a k_1$ (k_1 is a gain factor). Since the reference voltage V_a originates from the phase detector and is directly proportional to the phase difference $\Delta\phi$ between the reference and measured pulses,

increase in the load the motor slows down because the motor torque T_e is now too small. This immediately causes a progressive increase in the phase difference between ω_{ref} and ω, and hence an increase in V_a. This process continues until the voltage V_a is large enough to ensure a sufficiently large motor torque. The new equilibrium has then been reached.

To demonstrate the need for the damping by means of the frequency/voltage converter, we represent the circuit shown in the block diagram of fig. 3 in control-engineering terms (*fig. 4*). The appropriate transfer function is shown in each block; the individual transfer functions can be combined to give the transfer function G of the complete control loop. The independent variable in these functions is the complex frequency $s = \sigma + j\omega$, used as an operator in control and circuit theory. The same blocks as in fig. 3 are enclosed by chain-dotted and dashed lines. It can be shown from fig. 4 that for $k_3 = 0$ the transfer function for the open loop is given by

$$G = \frac{kk_1k_2}{Is^2(1 + s\tau_r)}, \qquad (4)$$

where τ_r is the time constant of the rotor circuit. A control loop with such a transfer function is always unstable when closed. If k_3 is made greater than zero the loop becomes stable. The transfer function of the open loop is then:

$$G = \frac{kk_1}{Is(1 + s\tau_r)} \frac{k_2(1 + k_3s/k_2)}{s}. \qquad (5)$$

For every combination of parameters there is a value of k_3 that will give critical damping of the transient effects.

Apart from the quantization that arises because reference and measured values are only compared at discrete times, the transfer function of this frequency-analog control system is identical with that of a conventional *PI* control for the speed. (In which the control is provided by a signal *P* proportional to the speed and a signal *I* proportional to the integral of the speed.)

Up to this point we have been concerned with maintaining the equality $\omega_{ref} = \omega$. However, at the moment when a stationary motor is switched on, $\omega = 0$ and $\omega_{ref} > 0$. The phase-lock loop is therefore not in equilibrium. It has to be brought into equilibrium by the synchronizing circuit *Sy* (fig. 3). This is essentially a frequency-comparison circuit; if $\omega_{ref} \neq \omega$ it produces the required acceleration or deceleration of the motor, and when ω_{ref} has become equal to ω, *Sy* switches over to the phase comparison described earlier.

The function of *Sy* can be carried out by various kinds of circuit. A very simple circuit that makes use of the phase detector described above (fig. 2) is a five-stage ring counter — a shift register with the output connected to the input — that can count both backwards and forwards (i.e. the pulses can progress through the shift register in either direction). The ring counter used here has automatic latching: in the final state '5' it suppresses all forward-travelling pulses and in the final state '1' it suppresses all backward-travelling pulses. The reference pulses travel forwards, the measured pulses backwards.

If $\omega_{ref} > \omega$, then on average there are more reference pulses per unit time than measured pulses and the counter is switched erratically back and forth between '4' and '5'. If $\omega_{ref} < \omega$, then it is switched erratically between states '1' and '2'. This switching back and forth can be used in both cases to make the control unit — e.g. a controlled rectifier — give either maximum acceleration of the motor (for '5', '4'), or maximum deceleration (for '1', '2'). When the two frequencies have been made equal by this synchronization process, a characteristic pattern results. Depending on the sign of the previous frequency deviation, either two forward pulses appear between two backward pulses (for switching between '1' and '2'), or two backward pulses appear between two forward ones (switching between '4' and '5'). When this indication is received the synchronizing circuit switches to the phase detector: forward and backward pulses now appear singly but alternately, which shows that the ring counter is switching between '2' and '3' or '3' and '4'. The phase difference between forward and backward pulses is now converted with the aid of the phase detector as described earlier into (say) a negative control voltage for braking and a positive one for acceleration [6].

We shall now end our description of this first example of a frequency-analog motor control with a short discussion of the frequency/voltage converter and the sensing device.

The most usual way of converting a frequency into a voltage is by deriving the average of the pulse frequencies with the aid of a lowpass filter, but this process gives too large a delay to permit effective damping. Only a very fast process making use of a hold circuit is of any use; the operation is shown in *fig. 5*. The leading (or trailing) edge of each measured pulse
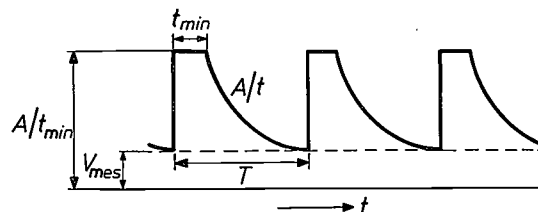
Fig. 5. Conversion of the measured frequency into a voltage V_{meas}. At a time t_{min} after a measured pulse a function generator starts to generate the function A/t (the time t is related to the leading or trailing edge of each pulse), and continues to operate until the next measured pulse arrives. The time between two pulses is equal to T, so that the voltage at that instant is A/T; this value is retained in a hold circuit, and since it is inversely proportional to T it is proportional to the repetition rate of the measured pulses.

starts an electronic function generator, which waits a constant time t_{min} and then generates a voltage curve of the shape $V(t) = A/t$; $V(t < t_{min}) = A/t_{min}$. On the arrival of the next measured pulse, at time $t = T$, the instantaneous value of the voltage is held in a hold circuit and the whole procedure starts again. In this way a measured value $V_{meas} = A/T$ can be obtained, where T is the period of the measured frequency. The application of this function generator thus enables a voltage proportional to ω to be derived for each pulse period.

Next we come to the sensor. Fig. 3 shows a well known type of sensor — a disc with equidistant slots. The disc is attached to the motor shaft; the frequency is obtained with the aid of a lamp and a photosensitive device connected to a pulse shaper. However, we have found it preferable to use a carrier-frequency technique, based on a new principle, in which single-sideband modulation is obtained from 'mechanical' signals. This can be done by using small rotary-field devices (synchros) driven by the drive shaft and with the carrier signal of frequency $\omega_0/2\pi$ supplied to the electrical

[5] See: E. M. H. Kamerbeek, Electric motors, this issue, p. 215.
[6] A system operating in a closely similar way but for another application and in only one direction has been described in detail by P. Diekmann in: Digitale Frequenz- und Phasenvergleichsschaltung für nachgesteuerte Oszillatoren, Int. elektron. Rdsch. 24, 231-232, 1970.

input (the stator or rotor). The signal obtained at the output then has either the sum frequency $(\omega_0 + \omega_r)/2\pi$ or the difference frequency $(\omega_0 - \omega_r)/2\pi$ depending on the sense of the mechanical rotation with respect to the rotary field.

This arrangement does not have the disadvantage of the slotted-disc sensor (and any other sensor that gives pulses at a repetition rate proportional to the shaft speed) that the scanning rate of such a discrete system is so low that this quantization limits performance and can introduce instability if there is a large decrease in speed. Our carrier-frequency technique also avoids another difficulty that can occur with the more usual double-sideband modulators. These produce two sidebands, $\omega_0 + \omega$ and $\omega_0 - \omega$; at low angular frequencies ω the two sidebands are very close and cannot be separated by using a linear filter.

Purely digital control

We now come to our second typical example of a frequency-analog speed-control system. The essential difference between this system and the first one is that here there is no analog transfer in the actual control loop (*fig. 6*), only switching. This approach makes the system very insensitive to interference.

As in fig. 3 the block *PD + Sy* (phase detector and synchronization) represents a bi-directional counter with almost identical control; but in this example of a single-quadrant drive — in which only one sense of rotation and motor operation are required — there are four states, not five, with latching in states '1' and '4'. The states of the counter have a different significance here. In the states '4' and '3' the amplifier *PA*, which amplifies the trigger pulses for the controlled rectifier, is gated off; while the counter has state '1' or '2' the thyristors Th_1 and Th_2 are triggered continuously. The reference pulses are derived from the zero crossings of the mains waveform by a pulse shaper PS_1: the mains voltage acts here as the reference signal. The measured pulses are derived from the zero crossings of the output voltage of the synchro generator by a pulse shaper PS_2. For the present we shall leave aside the functions of the phase modulator *PhM* connected in the signal path and the f/V converter for damping; their operation will be more fully explained later. The specified speed is presented here in the form of the carrier frequency for the synchro. The frequency range to be covered is given by

$$\omega_m \leqslant \omega_0 \leqslant \omega_m + 2\pi n_{max},$$

where ω_m is the angular frequency of the supply mains and n_{max} is the maximum speed of the motor (in revolutions per second). The angular frequency of the synchro output voltage is equal to $\omega_0 - \omega_r$.
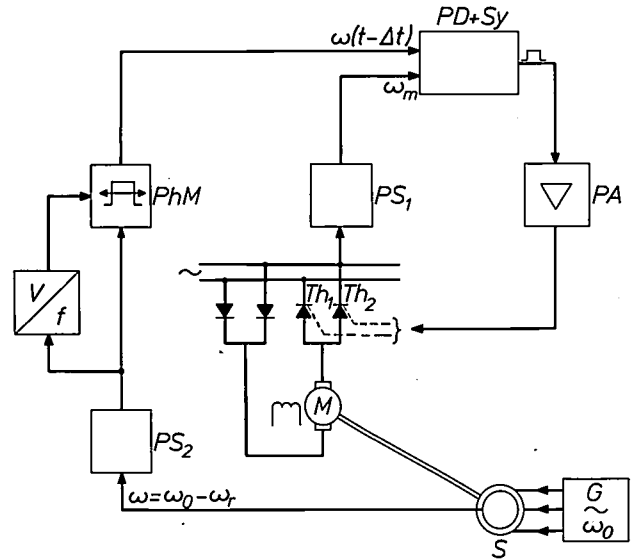
Fig. 6. Speed control for a d.c. motor. *M* motor. *S* synchro generator for generating the measured frequency. *G* three-phase generator. $PS_{1,2}$ pulse shapers. *PhM* phase modulator. *PA* pulse amplifier.

After the system has been switched on, $\omega_0 - \omega_r > \omega_m$, so that the ring counter switches erratically back and forth between states '1' and '2' because of the counting direction, as discussed earlier. Continuing with the convention adapted above the thyristors are then permanently triggered and the rotor voltage corresponds to the average of the complete half-cycle of the 50 Hz a.c. voltage. The maximum motor torque is then delivered and the motor consequently runs up to speed. This makes $\omega = \omega_0 - \omega_r$ smaller and when the situation $\omega_m = \omega$ is reached state '3' of the counter appears for the first time. Since state '3' is initiated by the reference pulse, then in our convention the trigger-pulse amplifier is gated off at the zero crossings of the mains voltage, so that triggering does not start at the beginning of the next half-cycle. Triggering does not start until the measured pulse switches the counter back to state '2'. From now on state '2' and state '3' are selected alternately, which has the effect of phase control of the mains voltage and also means that synchronization has been achieved. *Fig. 7a* shows the rotor voltage derived from the mains voltage, fig. 7*b* shows the output voltage of the synchro, fig. 7*c* the reference pulses derived from the mains voltage and fig. 7*d* the measured pulses derived from the synchro voltage. The hatching indicates the part of the cycle in which each thyristor conducts.

. If the load is increased the machine starts to run more slowly. Since $\omega = \omega_0 - \omega_r$, this results in a decrease in the phase difference between ω and ω_m and an increase in the firing angle α. The motor torque therefore increases and the control takes up a new

equilibrium. This is very similar to what happened in the first example for an increase in load, but one difference here is that the ring counter is involved in both measurement and control.

To explain the function of the phase modulator *PhM* and the frequency/voltage converter, let us look again at fig. 7. We can see at once that an intentional change in the phase difference between ω_m and ω will have an immediate effect on the motor torque because of the phase control. In the same way as in the first example damping can be provided by introducing a phase change proportional to ω. The block diagrams of the control are therefore identical for the two systems.

The phase modulator consists of a simple monostable circuit whose delay Δt can be controlled by a voltage. The control is taken from the output voltage of an f/V converter, like the one described earlier, via a highpass filter. The original value of the delay time of the monostable circuit, the 'operating point' for the control, has no steady-state significance since it merely produces a constant angular difference. It can in fact be used to apply angular corrections to a rotating motor with a control voltage. The dynamic operation is of course slightly affected by the small extra delay time.

Two other versions

To end this section we shall take a quick look at two frequency-analog control systems for induction motors, which are widely used in drive systems. The first system is for a slip-ring armature motor, the second for a squirrel-cage motor.

Since the *slip-ring armature motor* is in itself an electromechanical frequency-difference device, like a synchro, no sensor is necessary, and the frequency of the rotor current will serve as the measured quantity related to the speed (see *fig. 8*). To obtain the desired scanning speed the rotor-current frequency is multiplied by six, making use of the three rotor phases. In this way a measured-frequency range of 300-50 Hz corresponds to speeds of 0 to about 21 rev/s in a four-pole machine. This same frequency range also has to be covered by the reference generator *G* [7]. Control of the motor torque is obtained with the aid of a d.c. chopper *Ch* and a full-wave three-phase rectifier that converts the rotor current to d.c.; this d.c. current, which also flows through an external rotor resistance, is periodically interrupted by the chopper. The ratio of the time switched on to the time switched off (mark/space ratio) determines the effective resistance, giving rapid electronic control of the motor torque [5] [8]. As before damping is provided by phase modulation of the reference frequency.

The motor torque and speed of a *squirrel-cage motor* can be controlled by using an inverter; a voltage-controlled oscillator is then required that will control the frequency of the output voltage from the inverter, and also its amplitude, to obtain as constant a flux as possible in the machine; see *fig. 9*. The phase detector
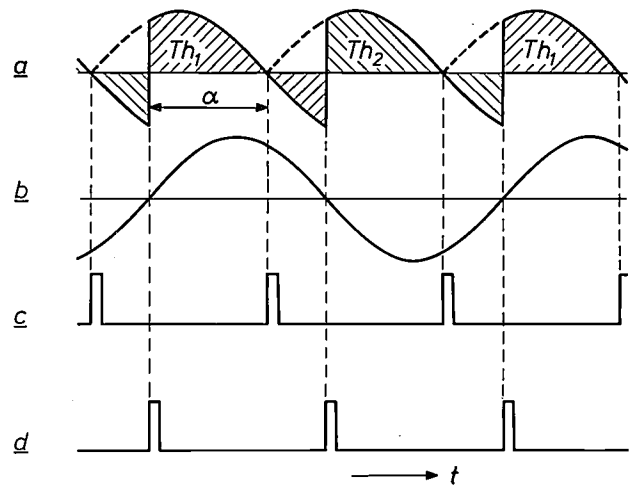
Fig. 7. *a*) Rotor voltage of the d.c. motor in the control system of fig. 6. At any given moment one of the two thyristors will be conducting (firing angle α); it will remain conducting after the mains voltage has changed sign since the energy stored in the magnetic field of the rotor still has to be discharged. This thyristor remains in the conducting state until the other one is triggered. *b*) Output voltage of the synchro. *c*) Reference pulses derived from the mains voltage. *d*) Measured pulses derived from the output voltage of the synchro; these determine the firing angle α of the thyristors.
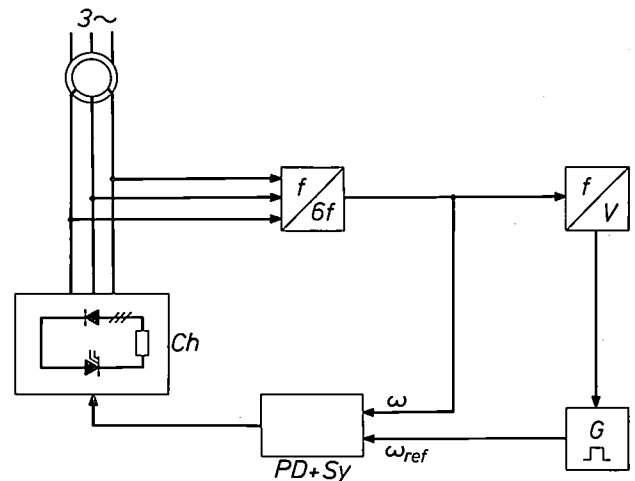
Fig. 8. Frequency-analog control of a slip-ring armature motor by measurement of the rotor frequency. This is multiplied by six and compared with the reference frequency produced by the generator *G*; the error signal resulting from this comparison controls a d.c. chopper *Ch*. This chopper interrupts the rotor current successively, after full-wave rectification in a three-phase rectifier, with a variable mark/space ratio and connects them to a resistive load. This varies the effective resistance in the rotor circuit and hence the motor torque.

and the synchronization stage operate in the same way as in the first example given in this section, while the measurement is made as in the second example, but with the aid of the upper sideband.

Since the flux in the machine is taken to be constant,

[7] J. Lemmrich, Frequenzanaloge Motorsteuerung mit kontaktlosen Bauelementen, Elektr. Ausrüstung 1965, No. 2, 58-61.

[8] H. Kalis and J. Lemmrich, D.C. chopper with high switching reliability and without limitation of the adjustable mark/space ratio, IEE Conf. Publn. No. 53, Part 1: Power thyristors and their applications, pp. 208-215, 1969.
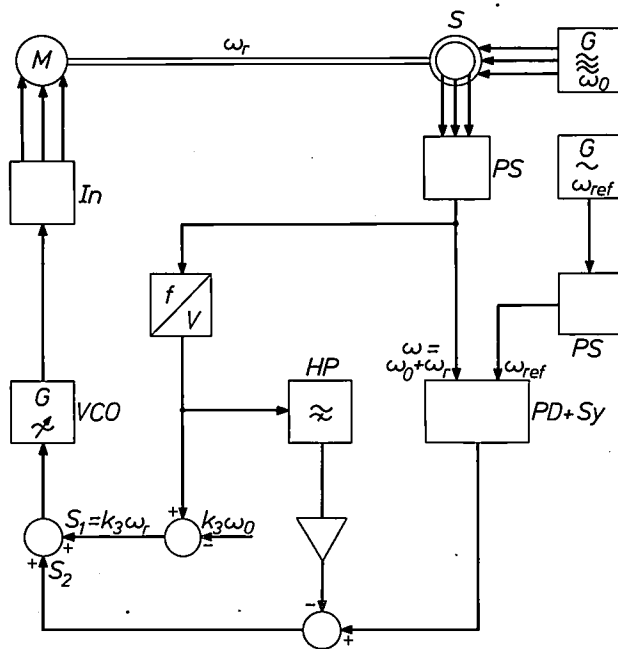
Fig. 9. Frequency-analog control of a squirrel-cage motor M by varying the frequency of the supply with the aid of an inverter In. The output frequency and amplitude of the inverter are determined by a voltage-controlled oscillator VCO. A synchro S measures the speed. HP highpass filter.

a larger motor torque can only be obtained at greater slip, because the torque of a squirrel-cage motor is approximately proportional to the slip in the operating region of the characteristic [5]. This proportionality requires an extra control of the supply frequency, which is obtained by adding to the control signal S_2 a signal $S_1 = k_3\omega_r$ proportional to the speed. The coefficient k_3 is given a value such that the rotating motor field is as close as possible to synchronism with the shaft at $S_2 = 0$. Damping is derived from the signal obtained through a highpass filter HP and the following amplifier.

Many more examples could be listed, such as the control of an induction motor by voltage chopping [9]. However, we hope that the explanations given above will allow such combinations to be worked out independently.

Advantages of frequency-analog control over conventional systems

Conventional systems of motor control are either completely analog or partly digital. A group that are completely analog are the PI and PID control systems; these use control signals proportional to the speed error (P), to its time integral (I) and in some cases to the time derivative (D). The most important advantages of frequency-analog controls over these analog systems are:

— Perfect angular precision, allowing accurate tracking.
— The angular position of drive shafts can be adjusted at will, even with the motor running, merely by changing the phase of reference or measured pulses.
— Much better signal/noise ratio (the information is not contained in the amplitude).
— Compatibility with electronic data-handling equipment.

Compared with high-quality digital control systems combined with a minor analog P-control loop, the advantages of frequency-analog control are:
— There is no need for the P-control tachogenerator.
— For the same discrimination of the sensing device the amplification that can be obtained in the control loop is much greater.
— The control quantity is continuously variable; this is not so in some conventional systems.

Very accurate angular tracking can be achieved by operating two frequency-analog control systems from the same reference generator, and constant speed ratios can be maintained by using frequency dividers and a common reference generator. Extremely low shaft speeds can be obtained, right down to zero. At zero speed the drive reacts like a torsion spring. With this arrangement tracking systems can therefore be run right up to speed synchronously, even from rest.

Drying-oven conveyor belt synchronized with a printing press

We now give a description of a frequency-analog control that we have developed for the conveyor belt of a drying oven. The frequency signals used not only provide the synchronization for the two drives, they are also applied for other control purposes.

Many commodities are packed nowadays in tins or in glass bottles or jars with closures, such as screw-caps or 'crown corks' made from tinplate. In the manufacture of the tins and closures sheets of tinplate about 1 m² in area are printed during one revolution of the cylinder of a printing press and then carried along between two carriers attached to the conveyor belt of a drying oven. *Fig. 10* shows a diagram of such a printing-and-drying unit. The drive for the conveyor belt has to meet the following requirements:
1. The conveyor belt and the printing press must operate in synchronism.
2. The speed of the belt should be continuously variable in a given range (e.g. 1 : 3).
3. The belt must never become so slack that the carrier

[9] A. Walraven, Controlling the speed of small induction motors by means of thyristors, Philips tech. Rev. 28, 1-12, 1967. K. Rennicke, Speed control of capacitor motors by varying the supply voltage, to appear in the next issue on electric motors.
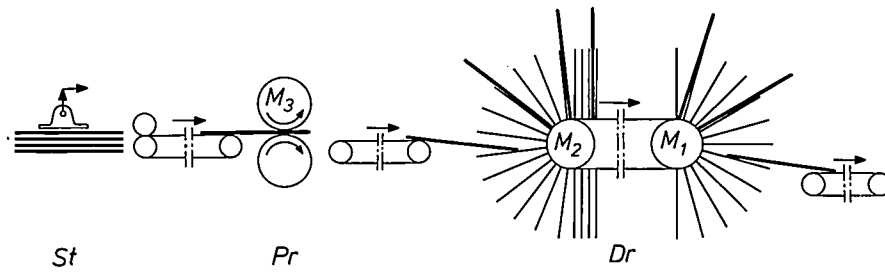
Fig. 10. The printing-and-drying unit for tinplate. St stack of sheets of tinplate to be printed. Pr printing press. Dr drying oven, with its conveyor belt and upright carriers to carry the printed tinplate through the oven. $M_{1,2,3}$ motors.

St Pr Dr

frames tip over. Also, the belt is so long that two driving motors are required.

4. The phase angle between the cylinder of the press and the separate carriers at the instant of transfer must be variable in steps or continuously, to enable different sizes of sheet to be loaded correctly into the drying section.

5. The time required to synchronize the conveyor belt to the selected phase must be less than 3 seconds. This is necessary for compatibility with other operations.

Previously all these requirements have been met by using mechanical devices such as shafts and clutches. However, there are limits to the successful use of such devices when the process is speeded up or the installation made larger. Electronic elements then have to be brought in. The solution to the problem is split into two parts: driving the conveyor belt to match a control value provided independently or by a sensor on the printing press, and the phase adjustment and synchronization at the correct phase angle.

Driving the conveyor belt

The conveyor belt is driven at each end by a controlled induction motor with gearing. Since the upper half of the belt is in tension, M_1 (fig. 10) has to supply the greater part of the total torque required to over-

come the friction, while M_2 determines the speed at the point where the tinplate sheets are transferred by supplying a greater or lesser additional torque. To obtain good dynamic operation and a satisfactory static stituation, it is found desirable to arrange that a constant torque, unaffected by the control and just too small to move the belt, is supplied by M_1; the additional torque required to operate the belt, and dependent on the load, is shared between the two motors. In this arrangement the ratio in which the additional torque is divided between the two machines is independent of speed and load, — a desirable condition for good dynamic operation.

The speed of the slip-ring armature motor M_2 is controlled by a frequency-analog system (*fig. 11*). The measured frequency is derived from the electrical rotor signal (as explained on p. 265). The reference frequency is produced in the rotor of a synchro S or by an electronic frequency generator G. Comparison of the reference and measured values in the block $PD + Sy$ gives a periodic switching function, which controls a chopper Ch that operates as an electronically controlled resistance [8]. *Fig. 12* shows examples of torque-speed characteristics.

Motor M_1 has a squirrel-cage rotor that makes use of the skin effect. The rotor is designed so that a voltage
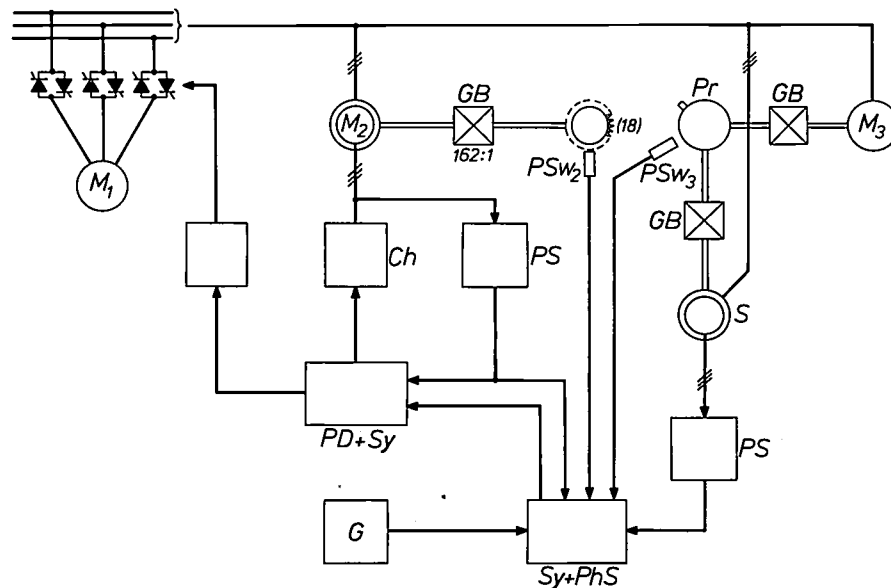
Fig. 11. Block diagram of the frequency-analog control for the conveyor belt. Motor M_1 provides the torque, motor M_2 determines the speed. M_3 driving motor for the printing cylinder Pr. GB gear-box. The block $PD + Sy$ controls the speed of the motor M_2 from the signals arriving via the chopper Ch and also provides a signal for the control of the torque of M_1. The synchronization circuit and phase-shifter $Sy + PhS$ synchronizes the belt with the printing cylinder and ensures the correct relative angular settings with the aid of the pulses from the proximity switches PSw_2 for the belt and PSw_3 for the cylinder. G generator for internal reference frequency.

of mains frequency induced in the stationary cage will only produce a current in the outer part of the cross-section of the conductor bars, because of the skin effect. This means that the stationary rotor has a relatively large effective resistance and hence a high starting torque. When the motor runs up to speed the frequency of the rotor current decreases, so that the skin effect also decreases and the effective rotor resistance becomes smaller. Because of this interaction the motor has a constant torque over a wide range of speeds; this independence of speed holds for a considerable range of stator-voltage variation, produced for example by phase control (see *fig. 13*).

The phase control for the torque motor M_1 is driven by a quasi-steady-state signal derived from the switching function δ of the speed motor M_2, which is proportional to the load (see figs. 11 and 12). This ensures that while changes in the load of the conveyor belt are sensed only by the frequency-analog control, they are shared between the two motors in an adjustable ratio. The torque distribution is made independent of the speed by using a multiplier circuit that takes the slip into account in generating the control signals.

Since the two motors M_1 and M_2 are both mechanically coupled by the belt and electrically coupled by the torque-sharing, precautions have to be taken to avoid instability. The dynamic characteristics can best be discussed with the aid of the control diagram of *fig. 14*. The frequency-analog control of M_2 has a transfer function G_{sc} (to a good approximation the same as for a *PI* control). The driving torque T_2 produced in M_2 first of all compensates the load T_{l2} and secondly it operates on the entire spring/mass system (G_1) of the conveyor belt; the angular velocity cannot therefore be derived directly by simple integration of the acceleration. At the same time a signal ε is derived from the control error of the frequency-analog control: this is the phase angle between reference and measured frequencies, and is represented by the mark/space ratio of the switching function; this signal ε determines the torque of M_1. This torque T_1 also operates both on the braking load T_{l1} and on the spring/mass system (G_3). The speed of each motor shaft is however still affected by the torque of the other motor through the spring/mass system $(G_2 = G_4)$.

The transfer functions G_1 to G_4 are determined by the interaction of the masses and springs in the conveyor belt, seen from the appropriate reference system. They are rational fractional third-degree functions of the complex frequency.

This system can easily be stabilized by putting the torque-control signal ε derived from the frequency-analog control through a lowpass filter of transfer function $1/(1 + s\tau)$, to attenuate the higher-frequency components responsible for the instability. An apparent disadvantage here is that the control of the torque in M_1 is then relatively slow compared with the speed of reaction of the frequency-analog control. However, since the complete system can be considered to be a slow system, cutting off the frequency response with a lowpass filter does give satisfactory stabilization. If faster operation were required for a different kind of installation, effective damping could for example be obtained with velocity feedback.
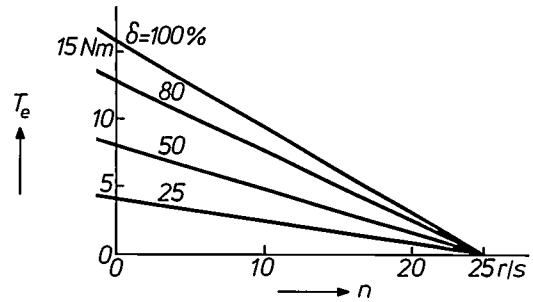
Fig. 12. The torque T_0 of the motor M_2 (fig. 11) as a function of the speed n; the rotor circuit is closed for a percentage δ of the time. The motor has four poles and is rated at 800 W.
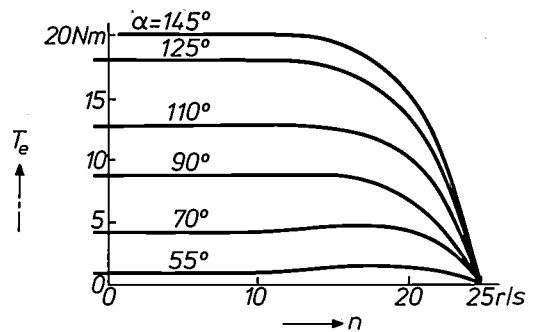
Fig. 13. Torque-speed characteristics of the squirrel-cage motor M_1 with skin effect, for phase control at various angles α. The torque is practically constant for a wide range of speeds.
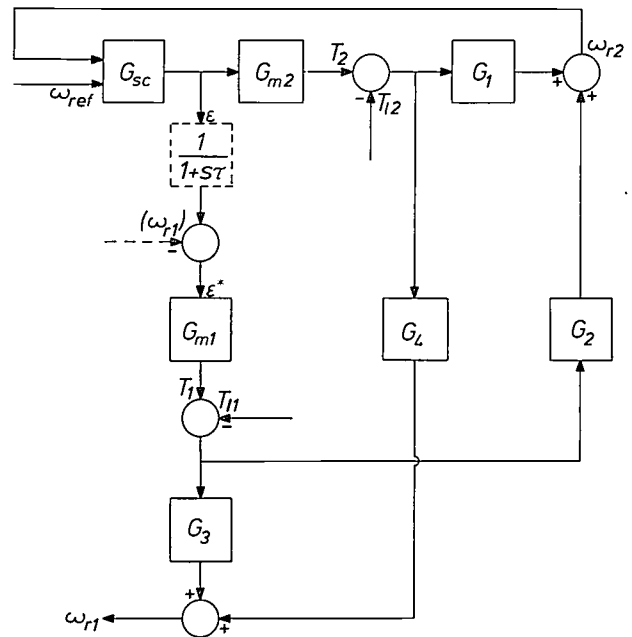
Fig. 14. Control block diagram of the conveyor-belt drive. G_{sc} control circuit. ε error signal. $G_{m1,2}$ motors $M_{1,2}$ with their control. G_1, G_3 driven spring/mass systems of the conveyor belt. $\omega_{r1,2}$ angular velocities of the motors $M_{1,2}$. Instability arising from the coupling between M_1 and M_2 through the belt can be damped out either by including a lowpass filter $1/(1 + s\tau)$ in the control of M_1 or by adding a signal of negative sign and proportional to the angular frequency ω_{r1}.

Phase adjustment and synchronization at the correct phase angle

For the conveyor belt to make the transition from the non-synchronized to the synchronized state, all that is required is to switch over from the internal reference frequency to the external frequency derived from the speed of the printing press. This would introduce a small transient because of the discontinuity in phase.

However, the situation becomes more difficult with the requirement that both drives must be synchronized with a particular phase angle between them, with no more than about 3 seconds delay between the start of the synchronization process and its completion. A condition for the fulfilment of this requirement is that there must be a difference in speed between the printing press and the dryer. There is a fixed relation between this difference and the time necessary for synchronization; the shorter this time has to be, the greater the difference in speed required, but the more marked the transient at the instant of synchronization. Some sort of monitoring device is also required that measures the relative angular position at the start of synchronization and compares it with the specified value. If the two values do not coincide, which is usually the case, then the change in the relative angular difference due to the difference in speed must be added to the value first measured until it reaches the specified value. Only then can the internal frequency be exchanged for the external one.

Although solutions to such problems have indeed been found, whether by using an analog method with servomotors, or a digital method whose pulse rate corresponds to the required setting accuracy for the angle, neither of these two kinds of sensing device are required in frequency-analog drives. Only simple pulse generators are required which indicate the separation in time between the successive carriers and between the separate revolutions of the printing cylinder. The fine subdivisions for accurate indication of the intermediate positions between two carriers are given by the measured frequency of the motor itself.

Starting from the frequency relation for the induction motor:

$$f_r = f_m - np,$$

where f_r is the frequency of the rotor currents, used as the measured frequency, f_m the mains frequency, n the number of revolutions per second and p the number of pole pairs, it can be shown that the displacement of the train of reference pulses results in a mechanical angular rotation of $360°/pa$ in one pulse interval T. Here a is a factor by which the measured frequency has to be multiplied. In the case described here a is equal to 6; this is because the zero crossings that occur in each of the three rotor phases are used in the multiplication process. For a four-pole machine and $a = 6$ the pulse interval therefore corresponds to an angular increment of 30°.

A displacement of the measured signal with respect to the reference signal of one pulse interval therefore corresponds to a 30° rotation of the motor shaft. If the motor is connected to gearing, then this corresponds to an angular difference of $30°/m_g$ at the output side, where m_g is the gear ratio. This increment of $30°/m_g$ indicates the fineness of the subdivision achieved by the frequency-analog measured signals, and hence the maximum setting accuracy. In the present case the gear ratio m_g is equal to 162, giving a subdivision into units of $30°/162 = 0.185°$. Such a degree of setting accuracy is by no means necessary, so that frequency dividers can be used to simplify monitoring and adjustment.

The definite relation between pulses and the angular positions of the motor shaft forms the basis of the circuit in which the measurement and adjustment of the phase take place. The composition of this circuit will now be described by examining the operations that take place in it.

After the drying-and-printing unit has been switched on a gate circuit is opened to pass the first pulse from the proximity switch PSw_3 for the printing cylinder, which releases an up-down counter C preceded by a frequency divider (see *fig. 15*). Two signals now appear at the input to the counter: the reference pulses from the rotor of the synchro driven by the printing roller and of frequency $f_{pr} = 6 (f_m - n_{pr})$, and the sixth harmonic of the mains $6f_m$ (in the expression for f_{pr} the quantity n_{pr} is the speed of the printing roller in revolutions per second). When the next pulse arrives from the proximity switch PSw_2 for the conveyor belt this counting process is stopped. The contents of the counter are now proportional to the number of revolu-
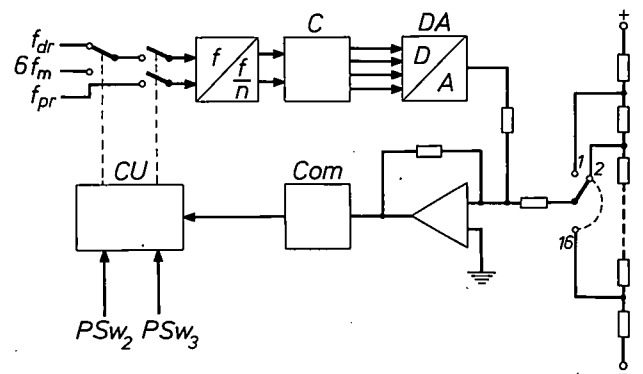
Fig. 15. Unit for measurement and control of the phase difference between printing cylinder and conveyor belt (fig. 11 *PhS*). *C* up-down counter. *DA* digital-to-analog converter. *Com* comparator. *CU* control unit. The phase difference is preset with the aid of the resistor network (*right*).

tions of the synchro and correspond accurately to the angle described by the printing cylinder during the measuring time. The proximity switch for the printing press supplies a pulse for each revolution of the printing cylinder, and the proximity switch for the conveyor belt supplies a single pulse each time a carrier passes (to give a phase angle of 20° between the separate carriers 18 teeth are required on the proximity-switch disc). The angular position of the printing cylinder with respect to the carriers therefore follows from the angle described by the printing cylinder between the two pulses.

Next, the counter input to which the sixth harmonic of the mains frequency is applied is instead switched to the measured frequency f_{dr} from the drying unit; this provides a continuous measurement of the difference in angle between the two machines. This relative angle varies continuously owing to the difference in speed. When the angular difference reaches 2π (corresponding to 108 pulses) the counter is set back to zero and started again because of the periodicity of 2π on rotation.

As soon as the angle — expressed by the state of the counter — between the cylinder and the belt, rotating at their different speeds, has become equal to the specified value, synchronization takes place as follows. The state of the counter is converted into an analog signal in a digital-to-analog converter (DA, fig. 15) and applied to the input of an operational amplifier. Currents that represent the specified value of the phase angle, but of the opposite polarity, are also supplied at the same input, so that comparison of these and the measured values is made right at the input of the amplifier. The control error is amplified and applied to a comparator Com that can distinguish between three states: positive error, negative error and equality (within a tolerance interval). Once the error has become equal to zero, the inputs of the counter are disconnected from the pulse generators and the internal reference frequency of the generator G (fig. 11) is exchanged for the external one from the printing unit.

In practice it is often necessary to alter the phase relation between printing and drying units while they are in operation. However, since the two units are already synchronized, there is no accelerating or decelerating reference frequency. This can then be derived from the reference frequency produced by the printing cylinder or from the measured frequency, with the aid of two special phase-shifters in which the applied pulses can be delayed as required.

The phase-shifters consist essentially of monostable circuits, connected in the path of the reference and measured pulses, with variable delay times. By increasing these delay times continuously the pulses can
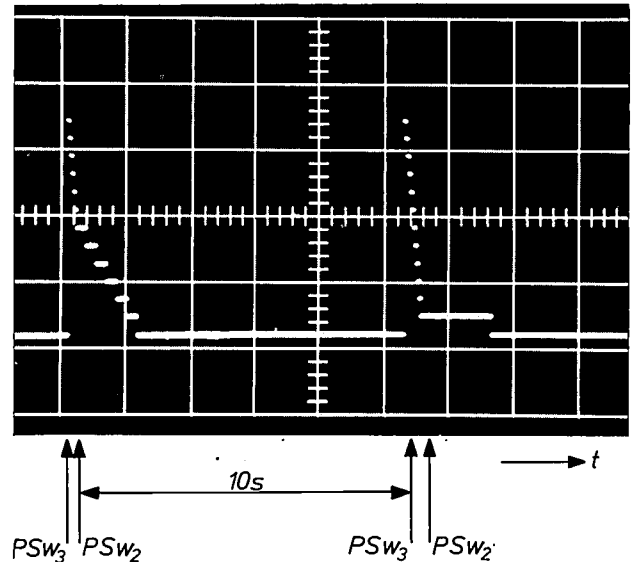
Fig. 16. The output voltage of the digital-to-analog converter at synchronization (*left*) and at the first phase check (*right*). After the pulse from the printing-unit proximity switch PSw_3 the counter is set back to zero (upper step); next the angular increments of the printing cylinder are added up in the counter (short steps) until a pulse arrives from the conveyor-belt proximity switch PSw_2. The number of increments indicates the phase angle of the cylinder with respect to the belt. From this moment on the two angular velocities are compared and only the differences in angular increments are counted in (long steps). The specified phase angle is reached in due course because of the differences in angular velocity. Synchronization takes place here; the counter remains at its final value. At the first check the phase angle is measured again; a phase correction of one angular increment is found necessary to correct a phase error that has arisen during the switchover.

be given a steadily increasing delay. This will cause the belt to speed up or slow down, depending on whether the variable delay is applied to the measured-frequency or reference-frequency path. The fundamental limit to the delay, and hence the phase shift, that is encountered when the next trigger pulse arrives before the previous delayed pulse has been delivered is avoided by setting the monostable circuits back to zero and suppressing the next delayed pulse to arrive.

These automatic phase-shifters come into action as soon as an angular error appears, due for example to a change in the specified phase. After the error has disappeared the phase-shifters switch themselves out of use automatically.

This system for phase measurement and control also includes a timer, which compares the value of the phase angle with the prescribed value every ten seconds. For this purpose the existing phase relation is first determined, as in the synchronization process. Any deviation from the specified relation is corrected by switching in the phase-shifters.

To illustrate the operation of the synchronization process, the output signal from the digital-to-analog converter is shown in *fig. 16*. The short 'steps' indicate

the measurement process for determining the angular difference at the time of the first pulse from the conveyor belt, the long 'steps' are angular corrections that are carried out before synchronization takes place. The first periodic phase check can be seen at the end of the trace after the first interval of 10 seconds. This check consists of a measurement and a correction of one step that had been lost because of the time taken by the synchronization process and now has to be made up.

Provided that the synchronization command does not coincide with switching on the complete printing-and-drying unit, but is only given when the machines are running at full speed, the synchronization time consists only of the correction time, which depends on the difference in speed. The delay arising between the synchronization command and the arrival of the first pulse from the printing unit and the measurement time that elapses between the appearance of the printing-unit pulse and that of the conveyor-belt pulse do not extend the synchronization time.

To obtain an estimate of the worst-case value of the synchronization time, we have to remember that this can be equal to the reciprocal of the difference in the number of revolutions per second of the printing cylinder and the belt. The reason for this is to be found in the periodicity of a rotation: at the instant when the synchronization command is applied the phase angle may have just overshot the specified value, so that an

extra angular difference of nearly 2π must be described before the specified value is reached. This means that with a printing-roller speed of 70 rev/min and a drying-unit speed of 100 rev/min the difference between the two is 30 rev/min or $\frac{1}{2}$ rev/s and the maximum synchronization time is therefore 2 seconds.

The specified phase angle is indicated by a switch that enables the specified value to be selected with the aid of a resistor network with 16 steps. Since the total spacing between carriers is 20° this means that the positioning accuracy is about 1.25°. Inside each of the steps there is a further uncertainty of $\pm$ 0.1°, which depends on the sequence in which the first measured and mains-frequency pulses happen to appear after the starting pulse for the measurement of the angle.

Summary. The well known phase-lock-loop technique from electronics can also be used with advantage in the speed control of electric motors. In this application the angular position of the motor shaft is continuously compared with the phase angle of a reference signal. The angular differences that arise are used to give a corresponding control of the motor, e.g. by phase control with thyristors. The phase comparison allows very accurate tracking to be obtained. The specified value, provided in the form of a frequency, can readily be transferred to a remote location and the digital data processing is not sensitive to interference. The principle can be applied with both d.c. motors and asynchronous motors; with asynchronous motors of the slip-ring armature type the rotor frequency provides the necessary measurement information for speed and phase control.

A tracking control for a drying and printing unit for tinplate sheets is described in detail as a practical example; a special feature of this system is that the relative phase can be altered when the unit is in operation.

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, 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 Forschungslaboratorium Aachen GmbH, Weißhausstraße, 51 Aachen,
   Germany   *A*

Philips Forschungslaboratorium Hamburg GmbH, 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, Netherlands.

W. Albers: Oriented eutectic crystallization.
Preparative methods in solid state chemistry, editor P. Hagenmuller, Academic Press, New York 1972, pp. 367-399.   *E*

W. Albers & L. van Hoof: Aligned eutectic thin film growth.
J. Crystal Growth **18**, 147-150, 1973 (No. 2).   *E*

G. A. A. Asselman, J. Mulder & R. J. Meijer: A high-performance radiator.
Proc. 7th Intersociety Energy Conversion Engng. Conf., San Diego 1972, pp. 865-874.   *E*

H. Bacchi & J. Denis: Electronique associée à des équipements de photographie ultra-rapide.
Acta Electronica **15**, 317-327, 1972 (No. 4).   *L*

C. Belouet: About the crystalline perfection of Nd-doped YAG single crystals.
J. Crystal Growth **15**, 188-194, 1972 (No. 3).   *L*

G. Bergmann: Überlegungen zur Ausnutzung der Emissionspolarisation für die polarisierte Kraftfahrzeug-Beleuchtung, I, II.
Lichttechnik **25**, 21-25 & 62-65, 1973 (Nos. 1 & 2).   *A*

M. Berth & C. Venger: Photodétecteurs rapides à état solide.
Acta Electronica **15**, 281-308, 1972 (No. 4).   *L*

J. W. M. Biesterbos & J. Hornstra: The crystal structure of the high-temperature, low-pressure form of Rh_2O_3.
J. less-common Met. **30**, 121-125, 1973 (No. 1).   *E*

R. Bleekrode & W. van Benthem: Resonance fluorescence of Mg atoms in the gas phase.
Philips Res. Repts. **28**, 130-132, 1973 (No. 2).   *E*

O. Boser: Influence of high concentrations of solute atoms on the critical flow stress of binary alloys, I. Theoretical foundations, II. Application to silver-, gold-, and copper-based alloys.
J. appl. Phys. **44**, 1033-1037 & 1038-1043, 1973 (No. 3).   *N*

H. Bouma (Institute for Perception Research, Eindhoven): Visual interference in the parafoveal recognition of initial and final letters of words.
Vision Res. **13**, 767-782, 1973 (No. 4).

J.-P. Boutot: Photomultiplicateur ultra-rapide à galette de microcanaux: le HR 300.
Acta Electronica **15**, 271-279, 1972 (No. 4).   *L*

M. J. G. Braken: Het fabriceren van schotelbalgelementen.
Lastechniek **39**, 4-6, 1973 (No. 1).   *E*

J. C. Brice & B. A. Joyce: The technology of semiconductor materials preparation.
Radio & Electronic Engr. **43**, 21-28, 1973 (No. 1/2). *M*

J. C. Brice & P. A. C. Whiffin: The platinum metals in crystal pulling. Crucibles and other apparatus.
Platinum Met. Rev. **17**, 46-51, 1973 (No. 2).   *M*

H. H. Brongersma & P. M. Mul: Analysis of the outermost atomic layer of a surface by low-energy ion scattering.
Surface Sci. **35**, 393-412, 1973.   *E*

E. Bruninx: The scintillation chamber: a high efficiency detector insensitive to source displacements.
Nucl. Instr. Meth. **106**, 613-614, 1973 (No. 3).   *E*

T. M. Bruton: The growth of single crystals by thermal diffusion.
J. Crystal Growth **18**, 269-272, 1973 (No. 3).   *M*

F. M. A. Carpay & W. A. Cense: Production of *in situ* composites by unidirectional crystalline decomposition of non-crystalline solids.
Nature phys. Sci. **241**, 19-20, 1973 (No. 105).   *E*

G. Clément, C. Delmare (C.E.A., Limeil) **& C. Loty:** Le Picoscope à microcanaux: un tube à balayage à haute résolution temporelle pour l'examen de phénomènes lumineux dans le domaine de la subnanoseconde.
Acta Electronica **15**, 369-371, 1972 (No. 4).   *L*

P. J. Courtois, F. Heymans & D. L. Parnas (Carnegie-Mellon University, Pittsburgh): Comments on 'A comparison of two synchronizing concepts by P. B. Hansen'.
Acta inform. **1**, 375-376, 1972 (No. 4).   *B*

P. A. van Dalen: Electric interaction between transverse electroacoustic waves in piezoelectrics of the Bleustein-Gulyaev geometry and a semiconductor.
Philips Res. Repts. **28**, 185-209, 1973 (No. 3).   *E*

M. Davio & J. J. Quisquater: Iterative universal logical modules.
Philips Res. Repts. **28**, 265-293, 1973 (No. 3).   *B*

M. Davio & A. Thayse: Representation of fuzzy functions.
Philips Res. Repts. **28**, 93-106, 1973 (No. 2).   *B*

M.-A. Deloron: Photodiode ultra-rapide.
Acta Electronica **15**, 265-270, 1972 (No. 4).   *L*

J. P. Deschamps: Parametric solutions of Boolean equations.
Discrete Math. **3**, 333-342, 1972 (No. 4).   *B*

J. P. Deschamps: Partially symmetric switching functions.
Philips Res. Repts. **28**, 245-264, 1973 (No. 3).   *B*

P. A. Devijver: The Bayesian distance. A new concept in statistical decision theory.
Proc. 1972 IEEE Conf. on Decision and Control and 11th Symp. on Adaptive Processes, New Orleans, pp. 543-544.   *B*

J. W. F. Dorleijn, W. F. Druyvesteyn, G. Bartels & W. Tolksdorf: Magnetic bubbles and stripe domains subjected to in-plane fields, I. Uniaxial anisotropy, II. Contribution of the cubic anisotropy.
Philips Res. Repts. **28**, 133-151 & 152-157, 1973 (No. 2).   *E, H*

J. Durieu & Y. Genin: On the existence of linear multistep formulas enjoying an 'h^2-process' property.
Philips Res. Repts. **28**, 120-129, 1973 (No. 2).   *B*

D. den Engelsen: Transmission ellipsometry and polarization spectrometry of thin layers.
J. phys. Chem. **76**, 3390-3397, 1972 (No. 23).   *E*

G. Engelsma: A possible role of divalent mangenese ions in the photoinduction of phenylalanine ammonia-lyase.
Plant Physiol. **50**, 599-602, 1972 (No. 5).   *E*

G. Eschard: Tube obturateur de type diode pour photographie ultra-rapide.
Acta Electronica **15**, 309-316, 1972 (No. 4).   *L*

L. J. M. Esser: Peristaltic charge-coupled device: a new type of charge-transfer device.
Electronics Letters **8**, 620-621, 1972 (No. 25).   *E*

W. G. Essers, G. Jelmorini & G. W. Tichelaar: Electrode phenomena with plasma-MIG welding.
Proc. 2nd Int. Conf. on Gas Discharges, London 1972, pp. 135-137.   *E*

E. Fabre: Bulk effects in photoluminescence measurements.
J. appl. Phys. **43**, 3788-3789, 1972 (No. 9).   *L*

D. G. J. Fanshawe: Sound field plotter.
SERT J. **7**, 37-38, 1973 (No. 2).   *M*

A. J. Fox: Longitudinal electro-optic effects in barium strontium niobate ($Ba_xSr_{1-x}Nb_2O_6$).
J. appl. Phys. **44**, 254-262, 1973 (No. 1).   *M*

G. Frens: Controlled nucleation for the regulation of the particle size in monodisperse gold suspensions.
Nature phys. Sci. **241**, 20-22, 1973 (No. 105).   *E*

C. J. Gerritsma & P. van Zanten: The dependence of the electric-field-induced cholesteric-nematic transition on the dielectric anisotropy.
Physics Letters **42A**, 127-128, 1972 (No. 2).   *E*

C. J. Gerritsma & P. van Zanten: An explanation of the observed field-induced blue shift of an imperfectly aligned planar cholesteric texture.
Physics Letters **42A**, 329-330, 1972 (No. 4).   *E*

J. Graf: L'intensificateur d'images à microcanaux. Applications à la photographie ultra-rapide.
Acta Electronica **15**, 357-362, 1972 (No. 4).   *L*

P. Guétin & G. Schréder: Tunneling spectroscopy and band-structure effects in *n* GaSb under pressure.
Phys. Rev. B **6**, 3816-3835, 1972 (No. 10).   *L*

F. E. L. ten Haaf (Philips Nuclear Applications Laboratory, Eindhoven): Colour quenching in liquid scintillation coincidence counters.
Liquid Scintillation Counting **2**, 39-48, 1972.

J. Hallais, C. Schemali & E. Fabre: Vapour epitaxial growth and characterization of $InAs_{1-x}P_x$.
J. Crystal Growth **17**, 173-182, 1972.   *L*

H. Haug & K. Weiss: Influence of a thin surface-dislocation layer on the Kapitza resistance.
Proc. 4th Int. Cryogenic Engng. Conf., Eindhoven 1972, pp. 129-130.   *E*

J. C. M. Henning: Exchange measurements by means of electron spin resonance and optical techniques.
Pulsed magnetic and optical resonance, Proc. Ampère Int. Summer School II, Baško polje 1971, pp. 179-205; 1972.   *E*

D. Hennings & G. Rosenstein: X-ray structure investigation of lanthanum modified lead titanate with A-site and B-site vacancies.
Mat. Res. Bull. **7**, 1505-1513, 1972 (No. 12).   *A*

K. Herff & E. Roeder: The investigation and measurement of thin layers.
Pract. Metallogr. **9**, 615-623, 1972 (No. 11).   *A*

W. A. H. J. Hermans (Philips Medical Systems Division, Eindhoven): On the instability of a translating gas bubble under the influence of a pressure step.
Thesis, Eindhoven 1973. (Philips Res. Repts. Suppl. 1973, No. 3.)

J. H. C. van Heuven & T. H. A. M. Vlek: Anisotropy in alumina substrates for microstrip circuits.
IEEE Trans. **MTT-20**, 775-777, 1972 (No. 11).   *E*

B. Hoekstra & R. P. van Stapele: Anomalous magnetic anisotropy and resonance linewidth in $CdCr_2S_4$.
Phys. Stat. sol. (b) **55**, 607-613, 1973 (No. 2).   *E*

J. P. Hulin: Parametric study of the optical storage effect in mixed liquid-crystal systems.
Appl. Phys. Letters **21**, 455-457, 1972 (No. 10).   *L*

B. A. Joyce: Some aspects of the surface behaviour of silicon.
Surface Sci. **35**, 1-7, 1973.   *M*

B. A. Joyce & J. H. Neave: Electron beam-adsorbate interactions on silicon surfaces.
Surface Sci. **34**, 401-419, 1973 (No. 2).   *M*

F. Kettel: The oxidation of zirconium at high temperatures.
Philips Res. Repts. **28**, 219-244, 1973 (No. 3).   *A*

E. Klotz & M. Kock: A simple ground-glass correlator.
Optics Comm. **6**, 391-393, 1972 (No. 4).   *H*

M. Kock & U. Tiemens: Tomosynthesis: a holographic method for variable depth display.
Optics Comm. **7**, 260-265, 1973 (No. 3).   *H*

A. J. R. de Kock: Microdefects in dislocation-free silicon crystals.
Thesis, Nijmegen 1973. (Philips Res. Repts. Suppl. 1973, No. 1.)   *E*

L. J. Koppens: Improved ferrite memory cores obtained by a new preparation technique.
IEEE Trans. **MAG-8**, 303-305, 1972 (No. 3).   *E*

E. Krätzig: Ultrasonic-surface-wave attenuation of gapless superconductors.
Phys. Rev. B **7**, 119-129, 1973 (No. 1).   *H*

J.-P. Krumme & P. Hansen: A new type of magnetic domain wall in nearly compensated Ga-substituted YIG.
Appl. Phys. Letters **22**, 312-314, 1973 (No. 7).   *H*

F. A. Kuijpers: Thermodynamic and magnetic properties of RCo_5 hydrides.
Ber. Bunsen-Ges. phys. Chemie **76**, 1220-1223, 1972 (No. 12).   *E*

F. A. Kuijpers: RCo_5-H and related systems.
Thesis, Delft 1973. (Philips Res. Repts. Suppl. 1973, No. 2.)   *E*

P. K. Larsen & S. Wittekoek: Photoconductivity and luminescence caused by band-band and by Cr^{3+} crystal field absorptions in $CdCr_2S_4$.
Phys. Rev. Letters **29**, 1597-1599, 1972 (No. 24).   *E*

J. H. J. Lorteye: Alfanumerieke displays met gasontladingen, lichtgevende diodes en vloeibare kristallen.
Ingenieur **84**, ET 145-150, 1972 (No. 50).   *E*

B. J. de Maagt & J. F. Verwey: Determination of the capture cross-section of hole-trapping centres in SiO_2.
Philips Res. Repts. **28**, 210-218, 1973 (No. 3).   *E*

H. H. van Mal & A. Mijnheer: Hydrogen refrigerator for the 20 K region with a $LaNi_5$-hydride thermal absorption compressor for hydrogen.
Proc. 4th Int. Cryogenic Engng. Conf., Eindhoven 1972, pp. 122-125.   *E*

A. P. J. Michels: C.V.S. test simulation of a 128-kW Stirling passenger car engine.
Proc. 7th Intersociety Energy Conversion Engng. Conf., San Diego 1972, pp. 875-886.   *E*

A. Mijnheer: Experiments on a two-stage Stirling cryogenerator with unbalanced regenerators.
Proc. 4th Int. Cryogenic Engng. Conf., Eindhoven 1972, pp. 83-86.   *E*

A. Mircea, A. Roussel & A. Mitonneau: $1/f$ noise: still a surface effect.
Physics Letters **41A**, 345-346, 1972 (No. 4).   *L*

B. J. Mulder: Preparation of thin single crystals of chalcosite and other cuprous sulphides.
Mat. Res. Bull. **7**, 1535-1542, 1972 (No. 12).   *E*

B. J. Mulder: Optical properties of an unusual form of thin chalcosite (Cu_2S) crystals.
Phys. Stat. sol. (a) **15**, 409-413, 1973 (No. 2).   *E*

F. L. van Nes (Institute for Perception Research, Eindhoven): Motor programming and errors in handwriting.
Neurophysiology studied in man, Proc. Symp. Paris 1971 (Int. Congress Series No. 253), pp. 420-425; 1973.

G. F. Neumark: Auger theory at defects — application to states with two bound particles in GaP.
Phys. Rev. B **7**, 3802-3810, 1973 (No. 8).   *N*

H. W. Newkirk: Observations on dislocations in tetraphenyltin and its isomorphs.
J. Chem. Soc. Dalton Trans. 1973, 12-14 (No. 1).   *A*

S. G. Nooteboom (Institute for Perception Research, Eindhoven): The perceptual reality of some prosodic durations.
J. Phonetics **1**, 25-45, 1973.

A. van Oostrom: The influence of temperature and field strength on the evaporation, desorption and emission from solid surfaces.
Ned. T. Vacuümtechniek **10**, 71-77, 1972 (No. 5).   *E*

G. den Ouden: Internal friction and ductility in the heat-affected zone.
Metal Constr. Brit. Welding J. **4**, 94-95, 1972 (No. 3).   *E*

L. J. van der Pauw: Impedance matrix of coupled piezoelectric resonators.
Philips Res. Repts. **28**, 158-178, 1973 (No. 2).   *E*

R. I. Pedroso & G. A. Domoto (Columbia University, New York): Perturbation solutions for spherical solidification of saturated liquids.
Trans. ASME C (J. Heat Transfer) **95**, 42-46, 1973 (No. 1).   *N*

G. Piétri: High-speed photoelectronics.
Acta Electronica **15**, 263-264, 1972 (No. 4). *(Also in French, pp. 261-262.)*   L

J. Polman & J. E. van der Werf: Cd(5^3P_1) and Cd(5^1P_1) atom density decay in a Cd-Ne afterglow plasma.
Physics Letters **42A**, 153-154, 1972 (No. 2).   E

H. Rau, T. R. N. Kutty & J. R. F. Guedes de Carvalho: High temperature saturated vapour pressure of sulphur and the estimation of its critical quantities.
J. chem. Thermodyn. **5**, 291-302, 1973 (No. 2).   A

F. Rondelez: Conductance measurements above a smectic $C \leftrightarrow$ nematic transition.
Solid State Comm. **11**, 1675-1678, 1972 (No. 12).   L

P. Röschmann: Compact YIG bandpass filter with finite-pole frequencies for applications in microwave integrated circuits.
IEEE Trans. MTT-21, 52-57, 1973 (No. 1).   H

J. A. J. Roufs (Institute for Perception Research, Eindhoven): Dynamic properties of vision, III. Twin flashes, single flashes and flicker fusion.
Vision Res. **13**, 309-323, 1973 (No. 2).

T. E. Rozzi & G. de Vrij: A series transformation for diaphragm-type discontinuities in waveguide.
IEEE Trans. MTT-20, 770-771, 1972 (No. 11).   E

U. J. Schmidt, E. Schröder (Elektro Spezial, Hamburg) **& W. Thust** (Elektro Spezial, Hamburg): Optimization procedures for digital light beam deflectors.
Appl. Optics **12**, 460-466, 1973 (No. 3).   H

A. A. Schüeli (Eidg. Techn. Hochschule, Zürich): Schnelle Parallel-Datenübertragung mit zeitbegrenzten Impulsen.
Thesis, Zürich 1973. (Philips Res. Repts. Suppl. 1973, No. 4.)

M. H. Seavey: Observation of light-induced anisotropy in ferric borate by acoustic resonance.
Solid State Comm. **12**, 49-52, 1973 (No. 1).   E

G. Simpson: Acoustic surface waves on paratellurite.
Electronics Letters **9**, 21-22, 1973 (No. 2).   M

A. Slob, H. A. van Essen & C. M. Hart: Aspecten van zakrekenmachines.
Informatie **15**, 1-16, 1973 (No. 1).   E

J. L. Sommerdijk: Influence of host lattice on the infra-red-excited visible luminescence in Yb^{3+}, Er^{3+}-doped fluorides.
J. Luminescence **6**, 61-67, 1973 (No. 1).   E

D. B. Spencer: A comparison of FM and digital modulation for direct television broadcasting from geostationary satellites.
E.B.U. Rev. tech. Part No. 137, 23-30, 1973.   M

W. T. Stacy & U. Enz: The characterization of magnetic bubble-domain materials with X-ray topography.
IEEE Trans. MAG-8, 268-272, 1972 (No. 3).   E

S. Strijbos: Burning-out of a carbonaceous residue from a porous body.
Chem. Engng. Sci. **28**, 205-213, 1973 (No. 1).   E

F. L. H. M. Stumpers, N. van Hurck & J. O. Voorman: Ein Empfänger für Zweiseitenband-AM und für Einseitenband-AM mit teilweise unterdrücktem Träger.
Rundfunktechn. Mitt. **16**, 202-206, 1972 (No. 5).   E

D. G. Taylor & P. Schagen: The application of channel image intensifiers to low light-level television.
Adv. in Electronics & Electron Phys. **33B**, 945-959, 1972.   M

A. Thayse: On some iteration properties of Boolean functions.
Philips Res. Repts. **28**, 107-119, 1973 (No. 2).   B

J. B. Theeten, J. L. Domange (E.N.S.C.P., Paris) **& J. P. Hurault:** LEED at 20 °K — a compulsory test of theory.
Solid State Comm. **11**, 1477-1479, 1972 (No. 10).   L

W. Tolksdorf: Herstellung hexagonaler Ferrit-Einkristalle mit Y-Struktur aus schmelzflüssiger Lösung.
J. Crystal Growth **18**, 57-60, 1973 (No. 1).   H

W. Tolksdorf, G. Bartels, G. P. Espinosa, P. Holst, D. Mateika & F. Welz: Controlled lattice constant mismatch by compositional changes in liquid phase epitaxially grown single crystal films of rare earth yttrium iron gallium garnets on gadolinium gallium garnet substrates.
J. Crystal Growth **17**, 322-328, 1972.   H

M. J. Underhill: The delay stabilised variable oscillator. A new type of stable tunable oscillator.
SERT J. **7**, 38-39, 1973 (No. 2).   M

H. Vantilborgh & A. van Lamsweerde: On an extension of Dijkstra's semaphore primitives.
Inform. Proc. Letters **1**, 181-186, 1972 (No. 5).   B

A. A. van der Veeke: High voltage pulse amplifier drives capacitive loads with short rise times.
Rev. sci. Instr. **43**, 1702-1703, 1972 (No. 11).   E

J. F. Verwey: Emitter avalanche currents in gated transistors.
Proc. 9th Annual Reliability Physics Symp., Las Vegas 1971, pp. 16-24.   E

J. Vlietstra & J. B. M. Lucassen (Philips Centre for Technology, Eindhoven): AEDBAR — a software system for designing and analysing planar bar mechanisms.
Software Pract. Exper. **3**, 29-42, 1973 (No. 1).

H. O. Voorman & A. Biesheuvel: An electronic gyrator.
IEEE J. SC-7, 469-474, 1972 (No. 6).   E

J. J. Vrakking & F. Meyer: Quantitative aspects of Auger electron spectroscopy; on the importance of attenuation of the primary electron beam.
Surface Sci. **35**, 34-41, 1973.   E

K. Walther: Sound velocity and magnetoelastic coupling in $CsMnF_3$.
Physics Letters **42A**, 315-316, 1972 (No. 4).   H

H. Weinerth: Integrated wideband amplifiers for VHF/UHF community antenna applications.
Electronic product news, Proc. EPN Seminar 'Practical design with integrated circuits', Paris 1972, pp. 135-171; 1973.   E

H. W. Werner: Instrumental aspects of secondary ion mass spectrometry and secondary ion imaging mass spectrometry.
Vacuum **22**, 613-617, 1972 (No. 11).   *E*

H. W. Werner & H. A. M. de Grefte: Investigation of surface layers by SIMS and SIIMS.
Surface Sci. **35**, 458-472, 1973.   *E*

G. Wesselink, D. de Mooy & M. J. C. van Gemert: Temperature determination of high-pressure optically thick gas discharges by a modified Bartels' method.
J. Physics D **6**, L 27-30, 1973 (No. 4).   *E*

P. W. Whipps, R. S. Cosier & K. L. Bye: Orthorhombic diglycine sulphate.
J. Mat. Sci. **7**, 1476-1477, 1972 (No. 12).   *M*

Field emission of electrons and ions

A. van Oostrom

Until recently the spectacular possibility of making the surface atoms of a specimen individually visible by means of a field-ion microscope was confined to metals possessing great cohesion, such as tungsten or platinum. The way in which this limitation was successfully removed is one of the subjects of the article below, which discusses the principles of field emission of both electrons and ions and also shows that the two methods lend themselves to both qualitative and quantitative investigations of the solid state.

Introduction

Electrons can overcome the potential barrier at the surface of a solid and be emitted if, for example, they absorb sufficient thermal energy (thermionic emission) or sufficient radiant energy (photoemission). It was originally thought that field emission, i.e. the emission of electrons which is caused by a strong electric field and which was first observed in metals towards the end of last century [1], could also be accounted for in a similar way. It was assumed that the applied electric field lowered the potential barrier at the surface sufficiently to make the emission of electrons possible, but this supposition proved to be incorrect. It was not until 1928 that R. H. Fowler and L. Nordheim [2] succeeded with the aid of quantum mechanics in giving a satisfactory explanation of the phenomenon, based on the 'tunnelling' of electrons through the potential barrier. Field electron emission, the first main subject in this article, forms the basis of field-emission microscopy, which will be discussed in greater detail below.

It was also in about 1930 that the possibility of ionizing hydrogen gas situated in a powerful electric field, i.e. field ionization, was predicted [3].

The field *emission* of ions, the second main subject of this article, starts with a field-ionization process in the immediate vicinity of the positive electrode. But whereas in field electron emission the electrons are liberated from the metal into the surrounding space, precisely the opposite takes place here: an electron from a gas molecule striking the metal tunnels its way

through the potential barrier at the surface to the metal. The electric field then causes the gas ion which has been left behind to leave the surface immediately. As in the case of field-electron emission, there is also a method of microscopy which is based on the field emission of ions. Because of the function which the gas plays in field-ion microscopy, it has been termed the 'image gas'.

Both forms of field emission were first used for microscopy by E. W. Müller [4]. This scientist has contributed so much in the area of field-emission and field-ion microscopy over a period of 35 years that he can rightfully be called the founder of both these investigation techniques.

The field-emission microscope, often abbreviated to FEM, is basically a very simple device [5]. One electrode is also the specimen, the other a fluorescent screen facing the specimen at a distance of several centimetres; both are situated in a vacuum chamber. The specimen must be tapered towards the screen and the tip radius must be so small that at voltages which can be conveniently used in practice (in the kilovolt range)

Dr A. van Oostrom is with Philips Research Laboratories, Eindhoven.

[1] R. W. Wood, Phys. Rev. (1) **5**, 1, 1897.
[2] R. H. Fowler and L. Nordheim, Proc. Roy. Soc. London A **119**, 173, 1928.
  L. Nordheim, Proc. Roy. Soc. London A **121**, 626, 1928.
[3] It was first found experimentally by E. W. Müller; see Z. Physik **131**, 136, 1951.
[4] For the first application of field-emission microscopy see E. W. Müller, Z. Physik **106**, 541, 1937; for the first application of field-ion microscopy see the article of note [3].
[5] A detailed description is given in: R. Gomer, Field emission and field ionization, Harvard Univ. Press, 1961.

the field set up is sufficiently strong to produce emission. The electrons emitted leave the specimen radially in straight lines and therefore produce a stereographic projection of the specimen tip on the screen (*fig. 1*). In practice, deviations from the spherical shape of the specimen tip and deviations from the radial field cause the paths of the electrons to be somewhat curved, resulting in a compression of the image. The magnification of the image is approximately equal to the distance of the sample from the screen divided by the radius of the specimen tip; it can easily be as high as a million. The resolving power of a field-emission
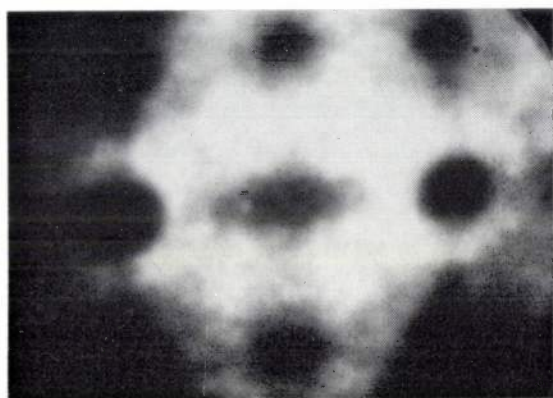
Fig. 1. Micrograph of a rhenium surface, made with a field-emission microscope at a voltage of several kV. The brightness variations correspond to differences in the work function of the emitted electrons.

microscope is limited to about 2 nm. This limit is set by the tangential velocity of the electrons and their de Broglie wavelength. An obvious way of overcoming this limitation is to use particles of larger mass (ions) and thus obtain both a lower tangential velocity and a lower de Broglie wavelength. In the field-ion microscope [6] (FIM) the resolving power is so high (0.2 nm) that the surface atoms of the specimen are imaged individually (*fig. 2*): it was the first instrument with which this was possible.

In the field-emission microscope an ultra-high vacuum (10^{-8} Pa, i.e. approximately 10^{-10} torr) is needed to ensure that the rate of adsorption of foreign gas molecules on the surface is low. The partial pressure of the contaminants may be higher in the FIM provided helium or neon is used as the image gas. In this case the field-strength at the surface of the specimen is so high that all other gas molecules are ionized before they can reach the surface. Vacuum technology has made such progress in the last twenty years that both the FEM and the FIM have been successfully developed into important tools for investigating solid surfaces.

Field emission can also be applied in scanning electron microscopy and in the mass-spectrometer investigation of organic substances. In the former applica-

tion a beam of electrons derived from a field-emission cathode is brought to an extremely fine focus and used to scan the object in a raster. The signal resulting from the scattering of electrons or the emission of true secondary electrons is used to modulate the intensity of another electron beam which describes a raster or pattern on a luminous screen in synchronism with the first beam. Individual heavy atoms can be imaged with this type of instrument [7]. With regard to the second of the two applications, mass spectra obtained with field-ion sources are found to be much simpler than those obtained with conventional ion sources, because the molecules under investigation disintegrate into fewer fragments.

Interest in field emission is also being shown in an entirely different field as a result of the discovery that an electrical discharge between two electrodes in a vacuum is generally initiated by the field emission of electrons [8]. This subject will be discussed briefly below.

To return now to our two field-emission microscopes, the FEM and FIM occupy a unique position among other aids and methods for studying surfaces because they enable differently oriented crystal planes of a specimen to be studied simultaneously. One of the ways in which these planes differ is in the density of the surface atoms, as reflected in differences in the work function of the electrons. By measuring the brightness variations of the FEM screen — or, as is usual, by measuring the electron flow in the different directions — the work functions of the different crystal planes of the specimen can be found. In the FIM the brightness differences on the screen are brought about by local field-strength differences caused by individual atoms. Most applications of the FIM are in metallurgy where it is used for example to study lattice imperfections and the way in which they originate or disappear. In particular the effect of bombardment with alpha particles, protons, other ions and neutrons and the resultant damage are being investigated in numerous laboratories. The simultaneous observation of the different crystal planes is also of interest in the study of adsorption, which is often very selective, especially in transition metals.

It is well known that small quantities of impurities can appreciably change the electrical properties of the surface of a solid. A hitherto much-used method of purifying surfaces consists in bombarding them with ions of inert gases and heating the specimen either simultaneously or subsequently. Such bombardment is very effective in removing impurities but also causes damage to the lattice, while it is also possible for inert-gas ions, depending on their energy, to penetrate a varying depth into the lattice (implantation). The

heating required to remove the implanted ions frequently leads to an equally undesirable segregation of impurities at the surface.

With the FEM and FIM another purification method in addition to those described above is possible. This consists in increasing the field-strength until the foreign atoms at the surface are 'pulled off' in the form of ions, a process known as field desorption. If the field-strength is increased still further, it is even possible to evaporate the actual material in ion form. This process, which is called field evaporation, results in surfaces which are completely free from lattice imperfections.
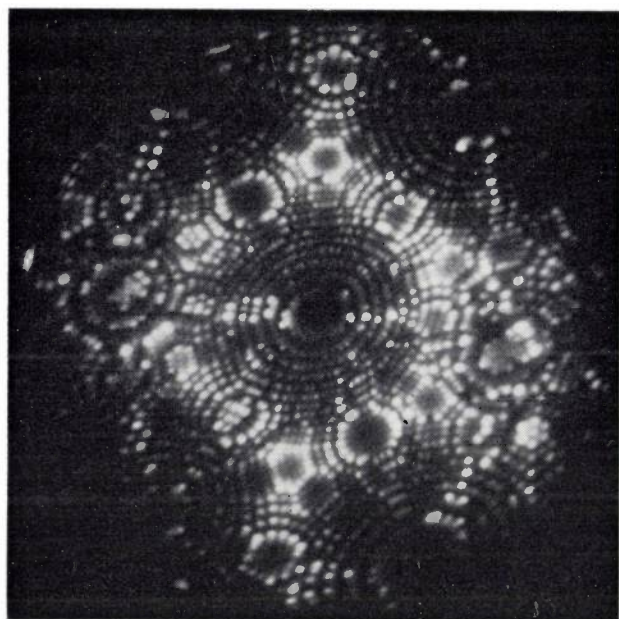
Fig. 2. Micrograph of a tungsten surface made with a field-ion microscope. Image gas helium. Temperature 78 K. The white spots are images of individual tungsten atoms. Every dark area corresponds to a crystal plane, of which only the peripheral atoms are imaged, thus forming a ring.

An important new tool for research into the solid state is a combination of an FIM and a mass spectrometer, known as an atom probe [9], in which the spectrometer is generally a time-of-flight mass spectrometer. With the atom probe atoms are freed from the specimen surface by field evaporation and identified one by one.

The intention in this article is to give an impression of the results that can be achieved with the FEM and FIM, placing special emphasis on results which yield an insight into adsorption phenomena. The better understanding now obtained of the nature of electrical discharges will also be briefly discussed.

So far as equipment is concerned, this article will deal chiefly with two recent improvements to the FIM [10]. Until recently the situation was that adequate brightness of the fluorescent screen could only be obtained with helium as the image gas. This meant

that only metals possessing great cohesion, such as tungsten and platinum, could be imaged without giving rise to field evaporation of the metal. By introducing a channel plate as an image intensifier we have succeeded in increasing the brightness of the screen to such an extent that neon or argon, for example, can be used as the image gas. This has prepared the way for more extensive application of the field-ion microscope. Not only tungsten and platinum but many other metals (fig. 3) can now be studied at lower field-strengths at which field evaporation does not occur.

Another improvement which the introduction of the channel plate has made possible is the use of colour to register small changes in the image [11], caused for example by field evaporation or surface migration. In one of the methods used by us a mixture of two phosphors is applied for the screen, generally Y_2O_2S-Eu (red) and $(Zn_{0.35}Cd_{0.65})$S-Ag (green). An exposure with a red filter is made before the change in the picture and another is made after the change with a green filter. The result is that the atoms which have remained in the same position are shown in yellow and the atoms which have disappeared in red, while the newly visible atoms will be green.

Field emission of electrons

Theoretical and experimental aspects

If at a metal surface there is an electric field directed at the surface, the potential barrier which an electron has to overcome on leaving the surface will have the form shown in *fig. 4*. The potential barrier is composed of two components: $-eEx$, the potential energy of the electron in the applied field, and $-e^2/4x$, the energy resulting from the action of the 'image force'. This force can be described as originating in a positive point charge generated by reflecting the electron in relation to the surface. From the relation for the potential barrier:

$$V = -eEx - e^2/4x,$$

it can easily be shown that the applied field lowers the level of the barrier by an amount $\sqrt{e^3E}$. This is sometimes referred to as the Schottky reduction due to the electric field.

[6] A detailed description is given in: E. W. Müller and Tien Tzou Tsong, Field ion microscopy, Elsevier, New York 1969.
[7] A. V. Crewe, J. Wall and J. Langmore, Science **168**, 1338, 1970.
[8] See for example the Proceedings of the Symposia Discharges and Electrical Insulation in Vacuum (Cambridge Mass. 1964, 1966; Paris 1968; Waterloo Ont. 1970; Poznan 1972).
[9] E. W. Müller, Naturwiss. **57**, 222, 1970.
[10] P. J. Turner, P. Cartwright, M. J. Southon, A. van Oostrom and B. W. Manley, J. sci. Instr. (J. Physics E), ser. 2, 2, 731, 1969.
[11] A. van Oostrom, Ned. T. Vacuümtechniek 9, 13, 1971 (in English).

The more powerful the electric field is, the lower, and also the narrower, the potential barrier will be; the two effects increase the tunnelling probability. If it is assumed that only free electrons can tunnel out of the metal, the electron flow at 0 K is given by [2]:

$$i = C_1 \frac{\beta^2 U^2 A}{\Phi} \exp\left[C_2 \frac{\Phi^{3/2}}{\beta U}\right], \qquad (1)$$

If the relation between the electron flow i and the applied voltage U is found experimentally for each crystal plane separately, the work-function difference between the various crystal planes can be calculated from the slope of the Fowler-Nordheim plots. Changes in the work function of a particular crystal plane resulting from adsorption can also be deduced from the variations in the slope. Measurements of this kind can
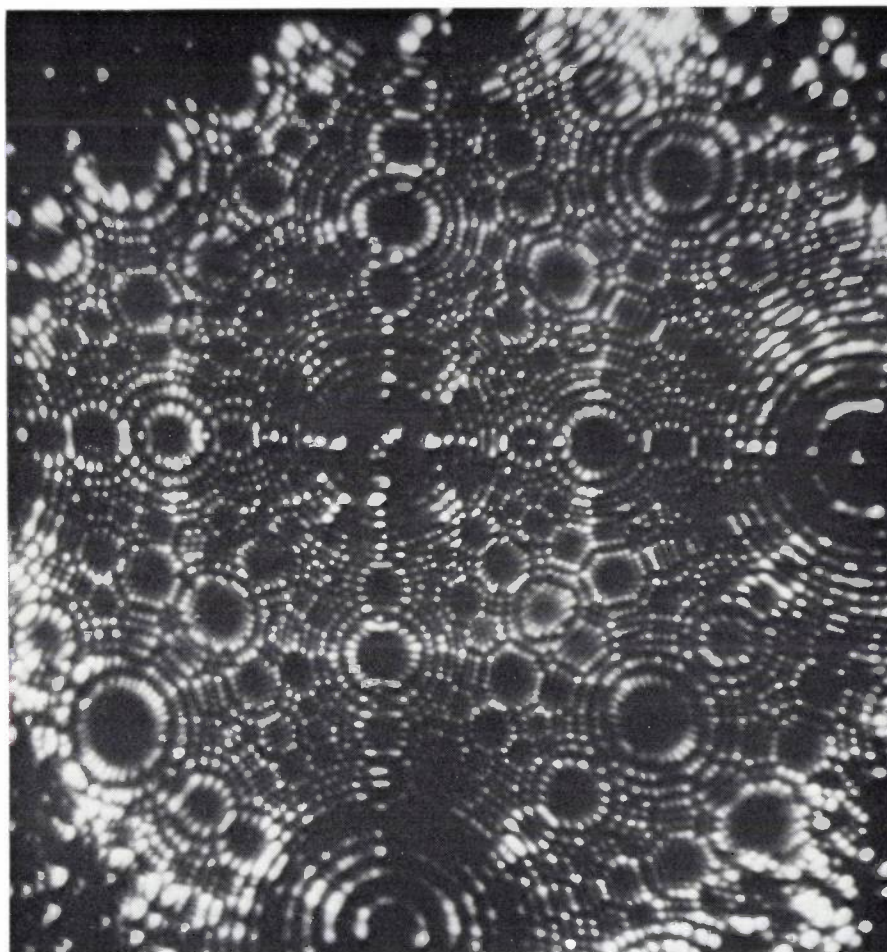
Fig. 3. Micrograph of a gold surface made with a field-ion microscope. Image gas neon. Temperature 20 K. (By courtesy of Dr M. J. Southon, Metallurgical Department, Cambridge University.)

where U is the applied voltage, β a geometrical factor, A the area of the emitting surface and Φ the work function. Although the image-force contribution means that factors C_1 and C_2 are not completely independent of U, a plot of $\log i/U^2$ versus $1/U$ is nevertheless a very good approximation to a straight line — the Fowler-Nordheim plot — whose slope is proportional to $\Phi^{3/2}/\beta$.

The energy distribution of the emitted electrons at a certain temperature and current density can also be calculated. The result of a calculation based, like the derivation of eq. (1), on the free-electron model is illustrated in fig. 5.

be carried out with an arrangement like that of fig. 6. The emitted electrons are incident upon an anode A containing a hole H behind which a collector $Coll$ is located. The enlarged image of the specimen is displayed on a phosphor layer Scr on the anode, thus enabling the equipment to be arranged in such a way that with magnetic or electrostatic deflection, or by rotation of the specimen, only the electrons from one particular crystal plane pass through the hole [12]. The electrons can also be passed through an energy analyser and their energy distribution determined. The full width at half maximum of this distribution depends on the work function and the applied electric field.
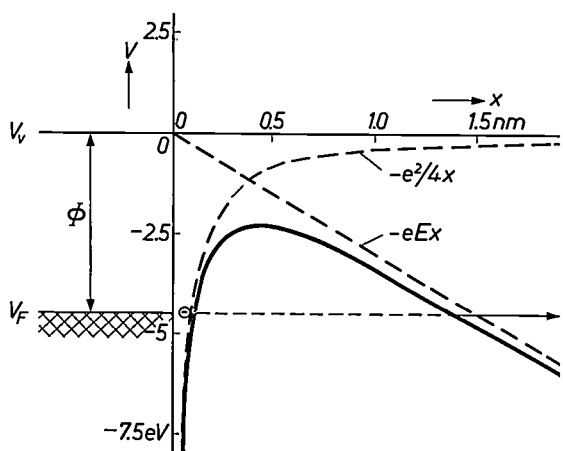
Fig. 4. The potential barrier 'seen' by an electron leaving a metal at an external field E of 3 GV/m and a temperature of 0 K. V potential energy of the electron. x distance from the metal surface. The contribution $-eEx$, the potential energy of the electron in a field of field-strength E, is indicated by a broken line, as also is the contribution $-e^2/4x$ made by the image force, i.e. the force exerted on the electron by the positive mirror-image charge. The figure also shows the Fermi level V_F of the electrons in the metal, the energy level V_v of the vacuum, and the work function Φ. The region with fully occupied electron levels is shown crosshatched. An arrow symbolizes tunnelling by the electron.
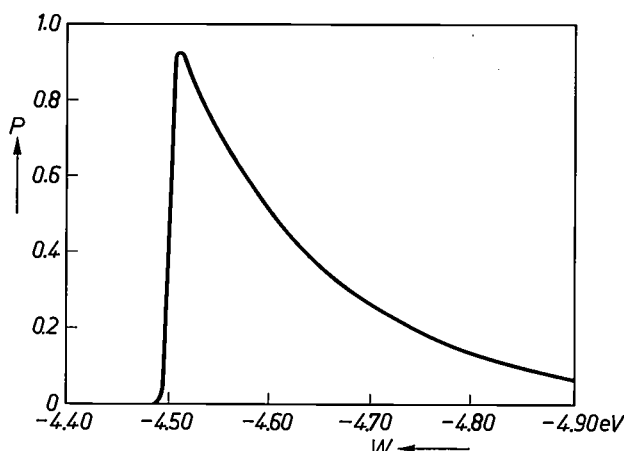
Fig. 5. Theoretical energy distribution (fraction P versus the total energy W) of electrons liberated from tungsten by field emission at a current density of 10^3 A/cm^2 and a temperature of 4.2 K. The calculation is based on the free-electron model and a Fermi level at -4.5 eV is assumed. The P values are normalized to unity for 0 K.
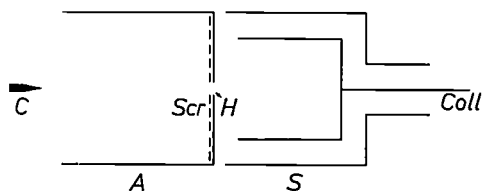
Fig 6. Schematic diagram of a field-emission microscope arranged for measurement on individual crystal planes. C cathode. A anode. Scr fluorescent screen. H hole in the anode. S shield. $Coll$ collector.

[12] A more detailed discussion is given in: A. van Oostrom, Thesis, Amsterdam 1965 (also published as Philips Res. Repts. Suppl. 1966, No. 1).

The theoretical and experimental results for the relation between the electron flow and the voltage and for the energy distribution of the emitted electrons are generally in fairly good agreement. Divergences are largely ascribed to deviations from the free-electron model, resulting for example from irregularities in the band structure or from the presence of surface states.

In the section which follows some results of our research into adsorption at the (100) plane of tungsten will be discussed: these results illustrate the potential of the method. The effect known as tunnel resonance will also be discussed.

Adsorption on the (100) plane of tungsten

The (100) plane of tungsten is of special interest to us because adsorption on it is often selective, and in other respects its behaviour also differs from that of most other crystal planes.

Atoms adsorbed on tungsten possess a dipole moment due to an interaction with the tungsten atoms at the surface. Depending on the donor or acceptor character of the adsorbed atom, this dipole moment will reduce or increase the work function. The former is the case when substances such as zirconium and the inert gases krypton and xenon are adsorbed (*Table I*), while the latter occurs upon adsorption of substances such as carbon monoxide and oxygen (at room temperature).

Table I. Some values measured by us for the reduction $\Delta\Phi$ of the work function which occurs as a result of adsorption on the (100) plane of tungsten. T is the temperature. The work function of pure tungsten is 4.65 eV. Its reduction causes the electron emission in the (100) direction to increase steeply in relation to other directions as a consequence of the great selectivity of the adsorption.

Adsorbate	T	$\Delta\Phi$
Rhenium	300 K	$-$ 0.05 eV
Zirconium	300	$-$ 1.95
Nitrogen	300	$-$ 0.31
Nitrogen	78	$-$ 0.51
Krypton	78	$-$ 0.69
Xenon	78	$-$ 1.53

The special character of the (100) plane of tungsten is indicated by the fact that the adsorption of nitrogen at room temperature causes a reduction of the work function which is not found upon adsorption of this gas on most other crystal planes.

Of the atoms which produce a marked reduction of the work function upon adsorption — and which therefore have a large dipole moment — zirconium adheres strongly and inert gases such as krypton and xenon weakly to the tungsten surface. Apparently, therefore, there is no direct relation between dipole moment and bond energy.

Some investigators have tried to explain the pronounced reduction of the work function by zirconium by the fact that the diameter of the zirconium atom coincides exactly with the shortest distance between two tungsten atoms in the (100) plane (0.316 nm). It seems to us, however, more appropriate to use this good fit to explain the high bond energy rather than the magnitude of the dipole moment. It then also becomes possible to explain the adsorption behaviour of rhenium which is also strongly bound but causes only a slight reduction of the work function. This atom combines a good fit and a low dipole moment.

Another respect in which the (100) plane of tungsten differs from the other crystal planes is that the measured value of the pre-exponential factor in eq. (1) is considerably larger than the theoretical value. This discrepancy disappears completely if a monolayer of nitrogen is adsorbed on the (100) plane at room temperature. This indicates the existence of surface states. The pre-exponential factors for other crystal planes are not greatly changed by adsorption.

More information on the subject can be obtained by analysing the energy distribution of the emitted electrons. It has been found that the distribution of electrons from the (100) plane of tungsten displays a prominent maximum at an energy level 0.35 eV below the Fermi level. Recent measurements have shown that this feature gradually disappears after adsorption of either zirconium, krypton or hydrogen [13]. This result is again an indication that surface states are present.

Tunnel resonance

It has just been shown that deviations from the free-electron model can disappear as a result of adsorption; conversely it is also possible for adsorption to give rise to deviations where none existed. This was first observed in experiments in which nitrogen was adsorbed on tungsten and then bombarded with electrons. At 78 K a new bond state was formed which led to marked divergences from the free-electron model [14].

When these divergences occur, they are often found to be associated with atoms or molecules for which the energy level of the valence electron is just below the Fermi level of the substrate. Quantum-mechanical resonance effects as illustrated in *fig. 7* then occur between the energy level of the valence electron and that of the metal. The resonance may cause the current density to increase by several orders of magnitude. This adsorption effect has mainly been observed in alkaline-earth metals.

The interaction with the metal electrons causes the level of the adsorbed atom to widen more pronouncedly the closer it is to the surface. The position and widening of the level can be found by measuring the energy distribution of the electrons.

When caesium, the heaviest of the alkali metals, is adsorbed on tungsten, tunnel resonance does not occur. This is connected with the fact that alkali metals have relatively wide bands in which the density of the energy levels is lower than in narrower bands. The number of energy levels that can lead to resonance is therefore relatively small. More success can consequently be expected with the alkaline-earth metals since they possess much narrower bands, especially for p and d electrons [15].

Field-electron emission and electric discharge

As was indicated in the introduction, the electrical discharge occurring between two flat electrodes in vacuum is also a subject of interest. It has been found, particularly for electrode spacings smaller than 1 cm, that the discharge is preceded by electric currents between the electrodes. These currents consist entirely of field-electron emission from the negative electrode and may initiate the electric discharge. The electron emission originates in whiskers or other sharp projections on the surface of the electrode. The electrons bombard a small area of the anode surface and will increase the temperature there. When the anode begins to evaporate, an 'anode discharge' is initiated. Because of the high current density in the whisker it is also possible that the latter will start evaporating sooner than the anode. In that case we speak of a 'cathode discharge'.

Because of the low thermal capacity of the emitting tip, it will reach an equilibrium temperature within several tens of nanoseconds after the voltage has been applied. The anode, on the
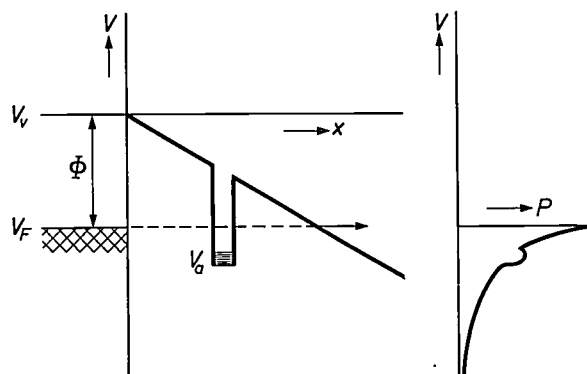
Fig. 7. Diagram to illustrate tunnel resonance resulting from adsorption during field emission of electrons. On the left the potential energy V of the electrons is plotted as a function of their distance x from the surface. The potential energy consists of a contribution from the external field and the energy V_a of the valence electron of the adsorbed atom. The contribution of the image force has been neglected (see fig. 4). The diagram on the right shows the energy spectrum of the emitted electrons (the fraction P versus the energy V). Exchange of electrons readily occurs between the energy level V_a, which has been widened by interaction with the metal surface, and the energy levels in the metal, which are situated in the same area. This means that electrons occupying the latter levels escape more easily than others, a fact reflected in a peak in the energy spectrum. V_F Fermi level. V_v vacuum level. Φ work function.
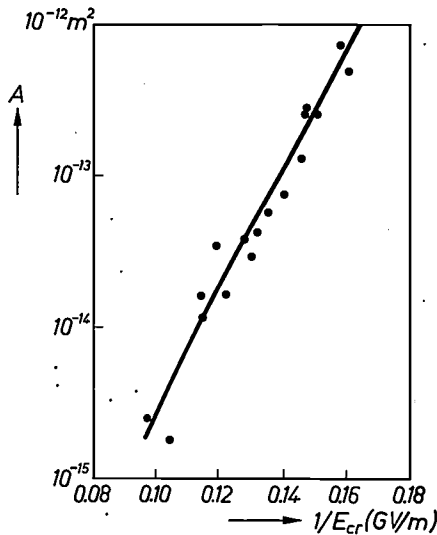
Fig. 8. The relation between the critical field-strength E_{cr} at which electrical discharge occurs and the surface area A of a number of emitting tungsten tips.
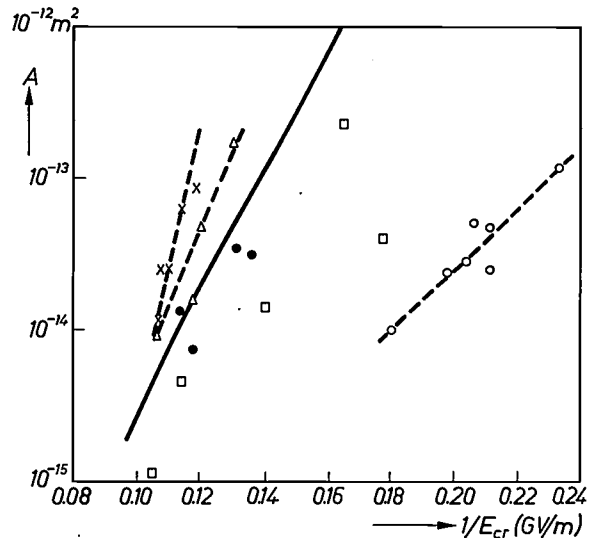

Fig. 9. Curves as in fig. 8 for tungsten which has been subjected to certain treatments: adsorption of oxygen (crosses), adsorption of carbon monoxide (triangles), adsorption of carbon (squares), argon-ion bombardment (white circles) and argon-ion bombardment followed by heating to 1200 °C (black circles). The continuous line is the curve for untreated tungsten, as in fig. 8.

other hand, is bombarded by a stream of electrons which is much less intensely localized, so that the anode only attains an equilibrium temperature milliseconds later. Advantage can be taken of this difference in thermal capacity to apply high-voltage pulses of several microseconds duration and cause a cathode discharge which will melt the emitting tip. An anode discharge obtained with a d.c. voltage causes appreciably more material to evaporate, which may have the undesirable effect of coating the insulators with a thin metal layer. The conditioning process using short pulses is therefore to be preferred to d.c. conditioning.

Discharge resulting from the evaporation of needle-shaped projections on the negative electrode occurs if the current density and hence the field-strength exceeds a certain limit. The critical current density is of the order of 10^{12}-10^{13} A/m², which corresponds in practice to a critical field-strength between 5 and 10 GV/m. The critical field-strength depends on the material and geometry of the emitting tip. In *fig. 8* the critical field-strength E_{cr} is plotted as a function of the emitting surface area A of the tungsten tips. Similar results have been obtained with rhenium. The value of the critical field-strength is also strongly influenced by quantities such as the surface purity. *Fig. 9* shows how the adsorption of oxygen and carbon monoxide, the presence of carbon as an impurity, and that of implanted argon ions can affect the relation between A and E_{cr}.

The equilibrium temperature reached in an emitting tip is chiefly determined by the Joule effect, thermal conductivity and the Nottingham effect. The Nottingham effect results from the difference between the mean energy of the emitted electrons and those which refill the holes formed in the Fermi sea by their emission (*fig. 10*). Heating occurs if the difference is negative, cooling if it is positive. For a particular field-strength
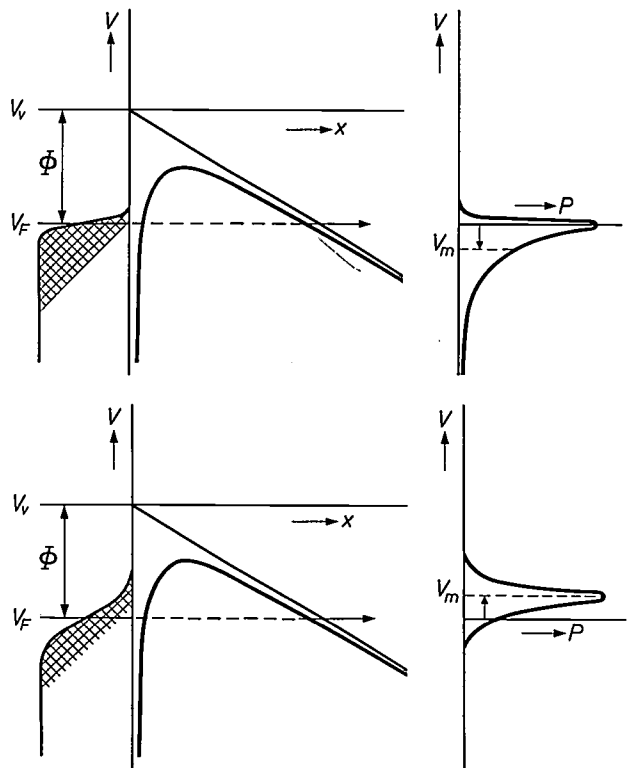
Fig. 10. Diagrams to illustrate the Nottingham effect. The two on the left show the occupancy of the energy levels in the metal, those on the right the energy spectrum of the emitted electrons (the fraction P versus the energy V). At low temperatures (*top*) the mean energy V_m of the emitted electrons is lower than the Fermi level and heating occurs as a result of replenishment of the electrons from the Fermi sea; if V_m is higher than V_F, as is the case at high temperatures (*bottom*), cooling occurs.

[13] C. Lea and R. Gomer, J. chem. Phys. **54**, 3349, 1971.
B. J. Waclawski and E. W. Plummer, Phys. Rev. Letters **29**, 783, 1972.
[14] W. Ermrich and A. van Oostrom, Solid State Comm. **5**, 471, 1967.
[15] E. W. Plummer and R. D. Young, Phys. Rev. B **1**, 2088, 1970.

and work function there is a critical temperature above which there is cooling and below which heating. For low-melting metals Nottingham heating is important, in fact more important than the Joule effect. In the case of high-melting metals the temperature equilibrium depends chiefly on Joule *heating* and Nottingham *cooling*.

Further research will undoubtedly lead to even greater control of the electrical discharge than our present knowledge, as represented by the results reported here.

Field-ion emission

Theory

The potential barrier which an electron has to pass through in the case of field-ion emission takes the form illustrated in *fig. 11*. As explained in the introduction, the direction in which the electron passes through the barrier is the opposite to that in field-electron emission: the electrons tunnel from the gas molecules through the barrier to the metal. Since the ionization energy of most gas molecules is appreciably higher than the work function of a metal surface (10·25 eV as compared to 4-5 eV), it means that the electric field required for field ionization must be stronger than for the field emission of electrons.

There are two requirements which must be met before field-ion emission can occur: not only must the potential barrier be sufficiently narrow and low, but the energy level of the valence electron to be removed must be higher than the Fermi level of the specimen — otherwise no tunnelling is possible. If the gas molecules are far from the surface of the specimen, the first condition is not satisfied; if they come too close to the surface, the energy level of the valence electron drops below the Fermi level and the second condition is not satisfied. There is therefore a narrow zone around the specimen tip in which ionization occurs. If the gas molecule has come so close to the specimen that the energy level of the valence electron has reached the Fermi level, the height and width of the potential barrier for the given field-strength have the lowest values at which tunnelling is still possible, and the ionization probability is greatest.

For a simple potential barrier of height $(\Phi - 2\sqrt{e^3E})$[16] and width $(I - \Phi)/eE$, the tunnelling probability is given by [6]:

$$D = \exp\left\{C_3(I - C_4E^{1/2})^{1/2}\left(\frac{I - \Phi}{E}\right)\right\}, \quad (2)$$

where I is the ionization energy of the gas molecule, Φ the local work function, E the electric field-strength and C_3 and C_4 represent constants.

If a situation occurs in which D has the same value for different image gases, the relation between the field-strengths can be derived from eq. (2). It is thus possible, for example, to compare the field-strengths at which field ionization is a maximum for various gases. The field-strength at this maximum can be determined for one particular gas by calibration with the aid of field-electron emission. In this process the polarity of the
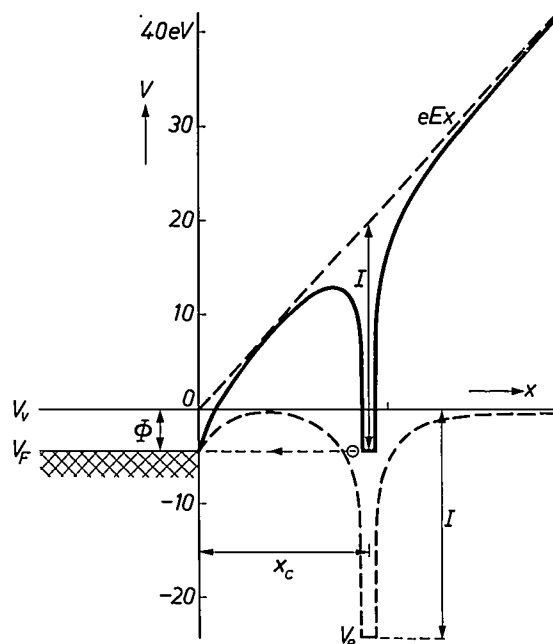
Fig. 11. The potential energy V of the valence electron of a helium atom in a field-ion microscope as a function of its distance x from the specimen surface, at a field-strength E of 44 GV/m and a temperature of 0 K. V_F Fermi level. V_v the energy level of the vacuum. A value of 4.5 eV is assumed for the work function Φ. I is the ionization energy of the gas. Broken lines indicate the contribution eEx — the potential energy of the electron in the applied field — and the potential energy of the electron without an external field. The latter contribution has a minimum V_e at the site of the gas atom. Unlike the situation in field electron emission, the electron tunnels towards the surface, and the ionized atom leaves the latter. The tunnelling probability increases as the potential barrier becomes narrower and lower. This does not apply, however, if the energy level of the electron drops below the Fermi level V_F, in which case tunnelling becomes impossible. The value x_c chosen for the distance of the atom from the surface in this figure is the one at which the two levels are the same. The ionization probability is then greatest.

voltage is reversed without altering the other conditions, and the gas is removed. A field-electron flow is then generated and the prevailing field-strength resulting from the intensity of this flow can be derived from the slope of the Fowler-Nordheim characteristic.

Table II lists a number of gases and the field-strengths at which field ionization is at a maximum, (a) as found theoretically from eq. (2) and (b) as measured by us. The agreement between theory and experiment can be considered excellent.

Measurement of the energy distribution of the ions has established that the ionization occurs in a zone which is only 20 pm (0.2 Å) wide. The number of molecules in this zone available for ionization is considerably larger than would be expected on the basis of the kinetic theory of gases. There is an extra supply of gas molecules since they are polarized in the inhomogeneous electric field. The equation for this supply Z_s is [17]:

$$Z_s \approx 4\pi r^2 \frac{p}{(2\pi MkT)^{1/2}} \left(\frac{\pi\alpha E^2}{2kT}\right)^{1/2}, \qquad (3)$$

where r represents the radius of curvature of the specimen, p the gas pressure, α and M the polarizability and the mass of the molecule, respectively, T the temperature of the gas and k the Boltzmann constant. As before, E is the electric field-strength. This relation will be used in one of the subsections which follow.

Field-ion microscope

For the proper functioning of a field-ion microscope it is essential that the specimen should be cooled. This is related primarily to the fact that the polarization of the gas molecules referred to above causes them to pass through the ionization zone at high velocity, thereby greatly reducing the probability of ionization. If, however, the specimen is cooled to lower temperatures, the velocity of the molecule after

Table II. Calculated (E_c) and measured (E_m) values of the field-strength at maximum field ionization for a number of gases, and the ionization energies of the same gases. Eq. (2) was used for the calculations.

Gas	I (eV)	E_m (GV/m)	E_c (GV/m)
He	24.5	44.0	44.0
Ne	21.5	34.0	34.5
Ar	15.7	19.0	19.0
H_2	15.6	22.0	18.8
N	14.5		16.5
N_2	15.5	18.5	18.6
CH_4	14.5	13.0	16.5

reflection and remaining a short time at the surface will have been greatly reduced, so that the ionization probability is much greater. In the second place, if the temperature of the image gas is low, the tangential velocity of the ions formed in the zone is low, which improves the resolving power of the field-ion microscope.

It is also important for the proper functioning of the FIM that the image gas should be correctly chosen. The factors determining this choice are as follows.

The first is the light output of the fluorescent screen. As was indicated in the introduction, we were able to increase this output by introducing a channel plate [18]. Until this was done the light output of the screen limited the choice to light gases such as hydrogen and helium.

The second factor is the radius of the ion. The smaller the radius, the higher the resolving power.

The third factor is the desirability of cooling the specimen. This limits the gases which can be used to those which do not condense at the cooling temperature.

The fourth factor is the level of the ionization energy. Helium possesses the highest ionization energy of any gas: 24.5 eV. This means that if helium is used as the image gas, the field-strength has to be so high that any other gases in the vacuum chamber are ionized before they reach the surface of the specimen. The choice of helium as the image gas ensures that the surface to be imaged remains completely free of adsorbed gas from the vacuum, even if the gas pressure is relatively high.

The fifth factor is the requirement that the field-strength necessary for field evaporation of the material being investigated should be higher than that at which the molecules of the image gas are ionized. It was this factor which previously restricted the application of the FIM using helium, with its high ionization energy, to the investigation of metals with great cohesion, such as tungsten. The ionization energy of the other gases which are now also employed, such as argon, xenon and methane, is lower than that of helium, which means that the field-strength used can also be much lower.

To enable the atoms of a crystal lattice to be imaged individually with the aid of a field-ion microscope, it is also necessary for the tip of the specimen to be practically spherical and 'atomically smooth'. This ideal shape is obtained with the aid of field evaporation, which ensures that the atoms located on sharp edges and projections are the first to be evaporated by the local increase of field-strength.

In the system which we employ, the specimen is cooled with a cryogenic liquid such as helium (4 K), hydrogen (20 K), neon (27 K) or nitrogen (78 K). There are also systems in which any desired temperature can be set, e.g. by flushing with cold gas. The image gas is admitted through a valve or by a diffusion process, e.g. through quartz in the case of helium.

Fig. 12 is a schematic diagram of a field-ion microscope with a channel plate. The gas ions produced by field ionization in the strong electric field at the surface

[16] The Schottky term derived earlier reappears here.
[17] M. J. Southon, Thesis, Cambridge 1963.
[18] J. Adams and B. W. Manley, Philips tech. Rev. **28**, 156, 1967.
        G. Eschard and R. Polaert, Philips tech. Rev. **30**, 252, 1969.

of the specimen *Sp* are accelerated radially towards the channel plate *Ch*, where they liberate electrons from the conducting layer at the channel-plate input. The number of electrons thus liberated depends on the mass and energy of the ion and also on the angle of incidence. The electrons entering the channels of the channel plate are multiplied by secondary emission. The emerging electrons are accelerated towards the fluorescent screen *Scr*. The total increase in the light output of the fluorescent screen may be as much as a factor of 10^5.

The higher luminous intensity of the picture obtained with the channel plate has enabled us to work not only with image gases other than helium but also to lower the gas pressure.

Some characteristic micrographs that were obtained with the field-ion microscope illustrated in fig. 12 will now be discussed.

Some characteristic micrographs

The interpretation of the images obtained with the field-ion microscope is often far from simple. This is not the case with the tungsten micrograph in *fig. 13*, which will be used to explain some characteristic features of FIM micrographs. The colour technique described in the introduction was used for this micrograph, which was made in order to study the process of field evaporation. Two exposures were made, one before the field evaporation using a red filter and one after it using a green filter. The object was imaged with helium as the image gas (with pressure 4×10^{-2} Pa, i.e. 3×10^{-4} torr) at a temperature of 78 K. The radius of curvature of the tungsten tip is approximately 15 nm, the magnification approximately three million times. Every spot in the image corresponds to an atom on the surface. Only the atoms which increase the field-strength locally are imaged. In practice these are the atoms at the edge of a crystal plane, resulting in a circular structure. The planes can be indexed by comparing the micrograph with the stereographic projection of a b.c.c. (body-centred cubic) lattice. The (110) plane in the middle of the picture and the four (211) planes surrounding it are characteristic. In the middle of the (110) plane there was a single atom which was removed by field evaporation and is therefore visible as a red dot. Some (red) atoms have also disappeared at other points on the surface as a result of field evaporation, while new (green) atoms have become visible. The photograph also shows a grain boundary cutting across the bottom of the pattern.

The colour technique also allows deformation effects to be studied. Plastic deformation can be caused by powerful electric fields, and may give rise to dislocations at the surface. Elastic deformation of the specimen as a whole can also occur. If the specimen is spherical and
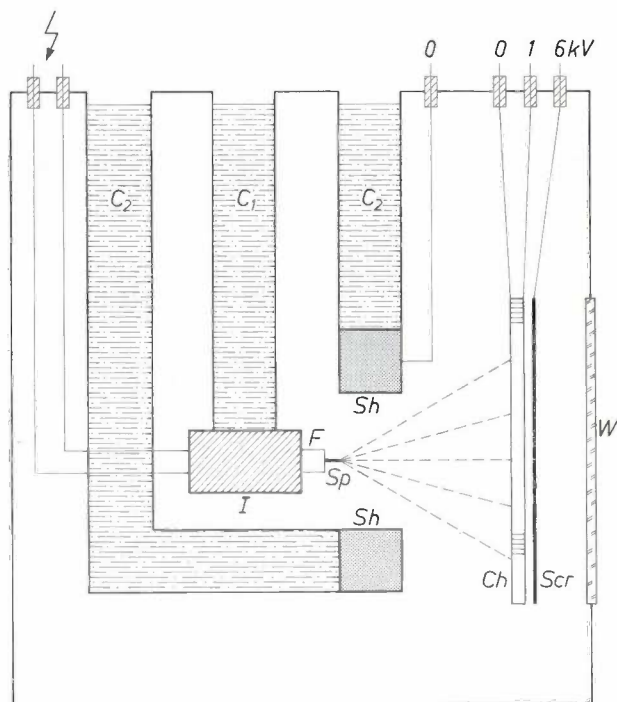
Fig. 12. Schematic diagram of a field-ion microscope incorporating a channel plate. *Sp* specimen. *F* filament enabling the specimen to be heated. *Scr* fluorescent screen. *Sh* shield electrode. *I* insulating block. *W* window. Cooling is effected with liquid neon in space C_1. To prevent rapid evaporation further cooling can be obtained with liquid nitrogen in spaces C_2.

isotropic, the deformation will not produce any change in the image and will consequently not be reflected in a shift of the post-deformation colour picture as compared to the first picture made. If, however, the tungsten tip is not spherical, a shift will take place. *Fig. 14* shows a tungsten tip with a grain boundary, imaged first with the aid of helium (green) and then with argon (red). To the left of the grain boundary the tip is approximately spherical and there the two pictures coincide accurately, as indicated particularly by the many yellow atoms in the (111) and (211) regions. To the right of the grain boundary, however, the material has been deformed as a result of the applied field, as can be seen from the divergence from the circular structure. Where the deformation is greatest, the helium and argon pictures do not coincide. The maximum shift here is approximately 0.3 nm.

The third micrograph to be discussed here (*fig. 15*) was prepared to study the effect of the image gas on the field evaporation. As stated, this process can be used to give the specimen tip the desired spherical shape. As a rule we carry out this process in the presence of the image gas, so that the specimen can be observed during the field evaporation.

It has been shown experimentally that molecules of

[19] E. W. Müller, S. B. McLane and J. A. Panitz, Surface Sci. **17**, 430, 1969.
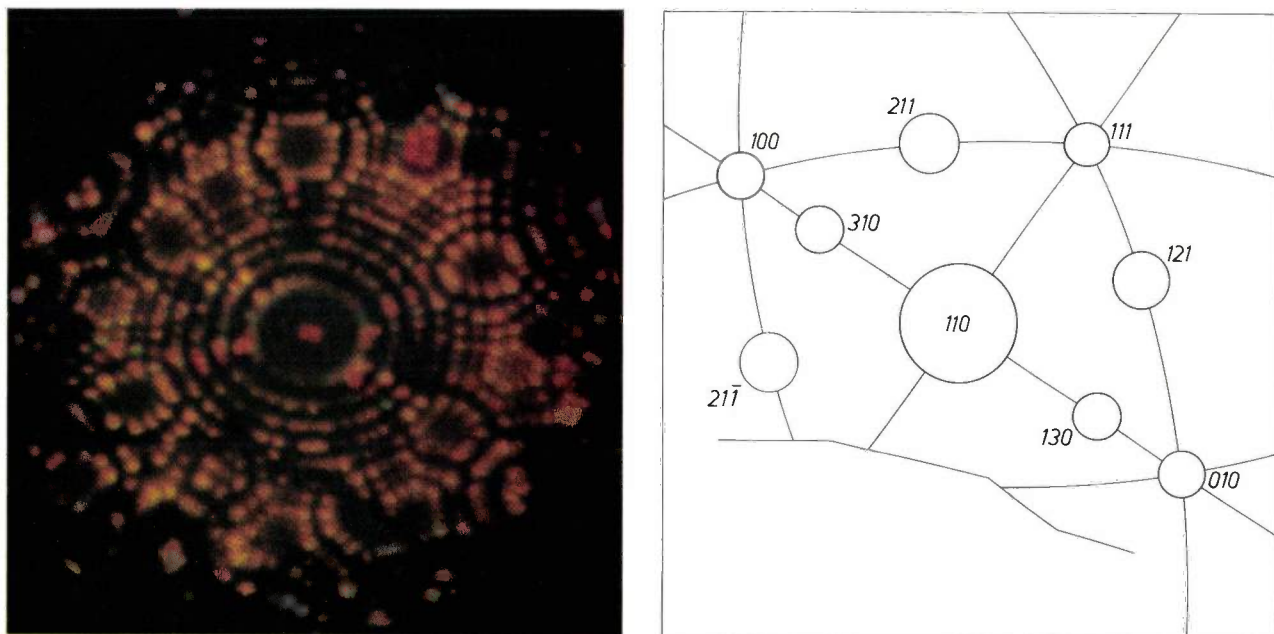
Fig. 13. Left, micrograph of a tungsten surface made with an FIM with 10^{-2} Pa helium and a voltage of 5360 V, illustrating the effect of field evaporation (at 6100 V). To enable the crystal planes to be indexed a stereographic projection corresponding to the micrograph has been drawn on the right. The fluorescent screen used consisted of red and green phosphors. One exposure was made with a red filter before field evaporation and another with a green filter afterwards. Where field evaporation produced no changes, the atoms are shown yellow. In the middle of the (110) plane there is a red spot where an atom obviously disappeared as a result of field evaporation. The effect of the field evaporation is also apparent at various other places in the form of red atoms (atoms which have disappeared) and green atoms (which have become visible). A grain boundary is visible at the bottom of the micrograph (shown on the right as an incomplete line).

the image gas (even at the imaging voltage) are adsorbed on the surface with a bond energy of $\frac{1}{2}\alpha E^2$ [19]. Although the polarizability α of inert gases is low, the high electric field-strength nevertheless makes adsorption possible. The presence of the adsorbed inert gases affects the field evaporation. As a general rule the
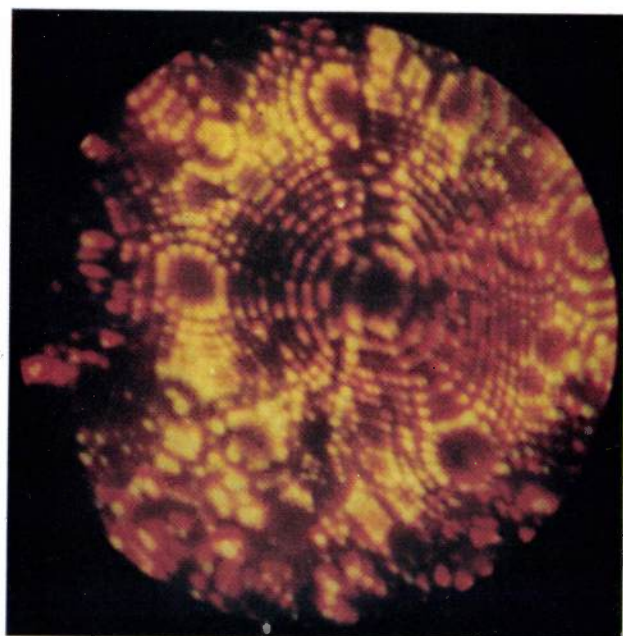
field-strength at which field evaporation occurs is lowered by the adsorbed gas. Fig. 15 clearly illustrates that the shape which the tungsten tip assumes as a result of evaporation in helium is different from that assumed in neon. The (110) area in the middle is flattened, and therefore not imaged, when evaporation takes place in neon (the field-strength is then too low because the curvature is too small). This flattening has also been observed in the (110) region in tungsten specimens with the (100) plane in the axial direction. The effect probably arises because inert gases are more strongly bound to the (110) plane than to other planes, and as a result of this interaction the bond energy of the tungsten atom is reduced.

Fig. 14. Micrograph of a tungsten surface made with an FIM, a first exposure having been made using helium as the image gas and with a red filter, followed by a second exposure with argon as the image gas and a green filter. The difference in imaging field-strengths between the two exposures (due to the different image gases) means that different electrostatic forces are exerted on the specimen. This has no effect on the micrograph at places where the specimen is perfectly spherical (to the left of the grain boundary) but it does have an effect where this is not the case (to the right of the grain boundary). Considerable deviations from circular symmetry are then observed and the red and green images do not completely coincide.

a   b
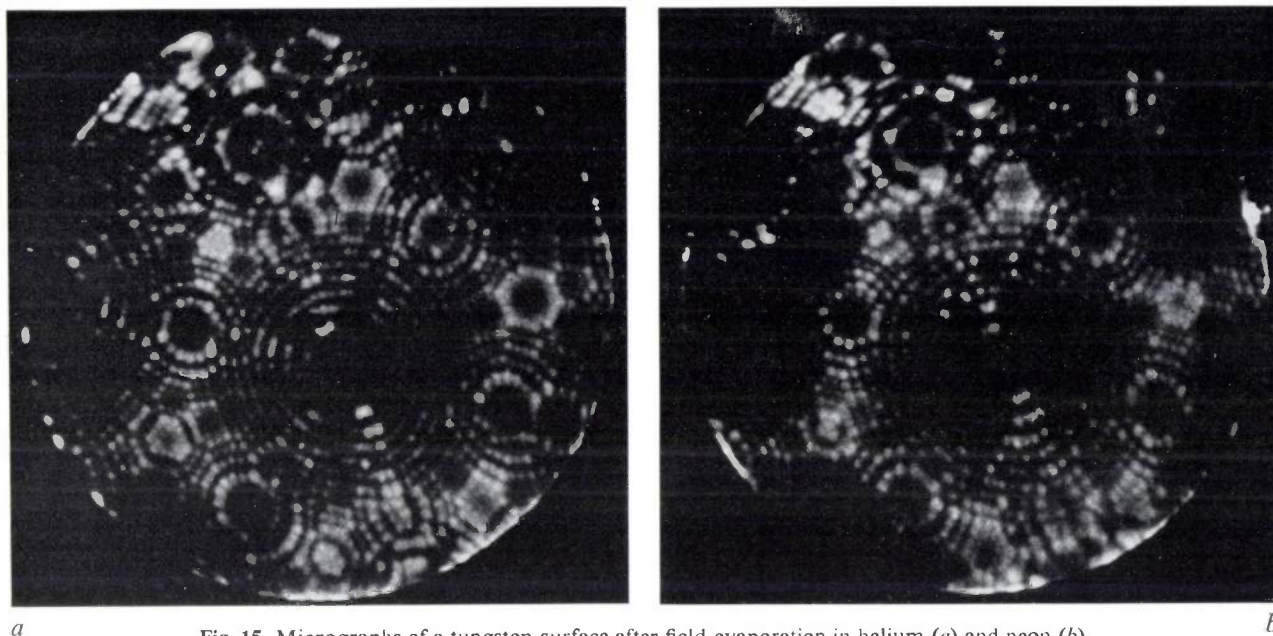
Fig. 15. Micrographs of a tungsten surface after field evaporation in helium (*a*) and neon (*b*). The (110) area (in the centre) is not visible in the latter case because of flattening caused by the field evaporation in neon. (This part of the specimen has too little curvature to give sufficient field-strength to allow imaging.)

Adsorption of nitrogen on tungsten

The remaining part of this article will be concerned with the adsorption of nitrogen on tungsten. The description of the process will show that the field-strength affects the interaction between the tungsten surface and the adsorbed nitrogen. If the field-strength is sufficiently high, the nitrogen causes a displacement of tungsten atoms which can be recorded with the aid of the colour technique described above. It will be shown at the same time that the field-strength also affects the probability of adsorption: the higher the field-strength, the larger the percentage of molecules prevented by 'premature' ionization from reaching the surface. The sites of the nitrogen molecules which do reach the surface and are adsorbed can be displayed indirectly with the aid of the colour technique, through the displacement of tungsten atoms referred to above. The number of molecules adsorbed can also be calculated. It follows from the agreement between the theoretical and the experimental data that it is possible to determine the number of nitrogen molecules adsorbed on the tungsten *quantitively* from an FIM micrograph.

An earlier study of the field desorption of nitrogen molecules which have been adsorbed on a tungsten surface showed that marked 'corrosion' of the surface occurs, the latter undergoing a complete reorientation. This behaviour has been ascribed to a reduction of the bond energy of the tungsten atoms as a result of the adsorption of nitrogen. A similar mechanism may

possibly operate in the case of field evaporation in the presence of hydrogen, a method frequently used to evaporate materials at a lower field-strength than is possible in vacuum.

Even without field desorption pronounced changes occur at the surface as a consequence of the electric field. It proved possible to establish this point in experiments in which the values of the electric field were such that the impossibility of field desorption could be safely assumed. In these experiments a tungsten tip was subjected to field evaporation in helium at 78 K and then imaged with 10^{-2} Pa argon at 19 GV/m. After the argon image had been recorded photographically, the argon was pumped from the microscope and nitrogen admitted, while the electric field was switched off to ensure that the nitrogen would be adsorbed. After the nitrogen had been evacuated to a pressure lower than 10^{-7} Pa, a micrograph was again made using argon.

This procedure was repeated a number of times, the field-strength being increased in steps after adsorption of nitrogen. After each step the image was again checked for changes with argon. Before the electric field-strength was increased by a further step, the argon was once more pumped off. We found that after a certain time, depending on the increase of the field-strength, the electric field resulted in a reorientation of a permanent nature. At 34 GV/m the reorientation of the surface was completed in a few seconds. *Fig. 16* is

a picture of the surface thus formed: the regular structure of a clean surface has completely disappeared.

From measurements with the atom probe it has been concluded that the imaged points are in fact derived from tungsten atoms and not from adsorbed nitrogen. The effect observed has been found to be independent of the adsorbed quantity of nitrogen over a wide range. The amount in our experiments was varied from 1 to 1500×10^{-4} Pa s.

The influence of the image gas was also investigated by repeating the above experiment at various values of argon pressure between 10^{-2} and 10^{-5} Pa. At higher pressures the corrosion was considerably accelerated by electron bombardment or perhaps as a result of energy dissipation by incident high-energy polarized argon atoms. After the field evaporation of two atomic layers the perfectly ordered surface structure was once more obtained.

In order to eliminate the influence of the 'foreign' image gas, the experiment was repeated with nitrogen itself as the image gas at a field-strength of 18.5 GV/m. *Fig. 17* shows a micrograph made with nitrogen, at a temperature of 78 K. It is obvious that in this case no reorientation has yet occurred although the image was obtained at a nitrogen pressure of 7×10^{-3} Pa. At such a pressure it would at first glance be expected that the surface would be rapidly covered with a monolayer of nitrogen since the probability of nitrogen adhering to tungsten at 78 K is practically unity. A number of the incident nitrogen molecules, however, will be ionized before they have reached the surface. The question is what percentage of the incident molecules can in fact reach the surface.

A good method of investigating this point is to take advantage of a unique feature of the field-ion microscope: protection of the surface by the electric field (field protection). If the surface is imaged with the aid of neon at 34 GV/m, there is no other gas, with the exception of helium, which can reach the surface. All molecules of any other gas will be ionized some distance from the surface. The molecular flow to the surface, and hence the adsorption, can be controlled by setting the field-strength to less high values.

This screening effect of the electric field is illustrated by the two micrographs in *fig. 18*. Both exposures are of the same tungsten specimen at 78 K and neon pressure of 10^{-2} Pa. The image voltage is 4600 V. The microscope also contains nitrogen with a partial pressure of 2×10^{-5} Pa. All other partial pressures are of the order of 10^{-8} Pa or lower. The micrograph on the left shows the effect of reducing the field-strength to zero for 10 seconds, and the one on the right that of reducing the field-strength to 23 GV/m for the same period. In both cases the specimen is shown in red before the field
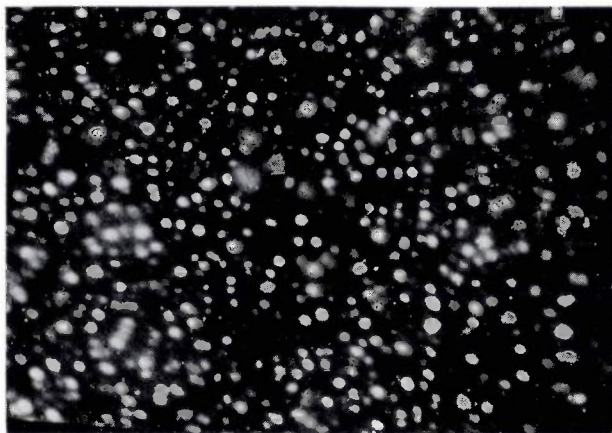
Fig. 16. Micrograph of a tungsten surface on which nitrogen has first been adsorbed and which has then been subjected to a field-strength of 34 GV/m (image gas argon, temperature 78 K). A reorientation of the surface then takes place within several seconds, the neat ordered structure of the clean surface disappearing entirely.
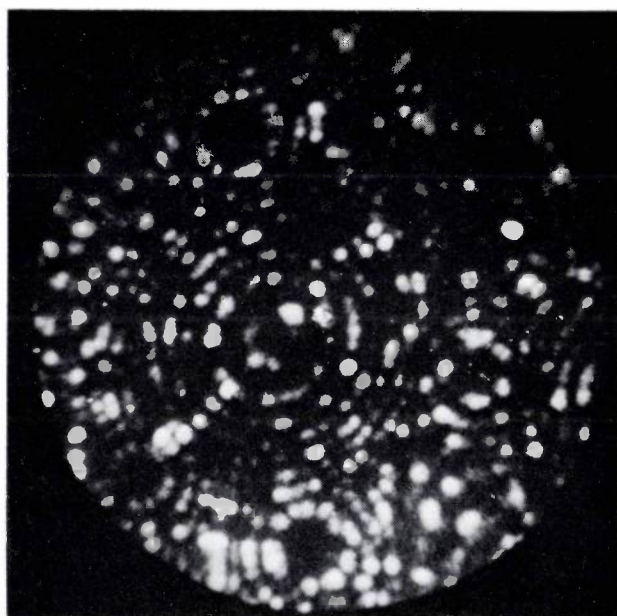
Fig. 17. Micrograph of a tungsten surface at a field-strength of 18.5 GV/m, using nitrogen as the image gas (6.5×10^{-3} Pa). Temperature 78 K. Reorientation has not occurred although in view of the effect of nitrogen adsorption demonstrated in fig. 16 it would have been expected.

reduction and in green after the field reduction. Changes have occurred in the photograph on the left, as can be seen from the red and green atoms; adsorption was apparently possible here. In the photograph on the right there has been no adsorption: all atoms in the picture are yellow.

The number of molecules that can reach the surface depends on the gas pressure as well as on the probability of ionization. Polarization will cause this number to increase in proportion to the field-strength until field ionization makes it fall sharply again. This number can be easily calculated.
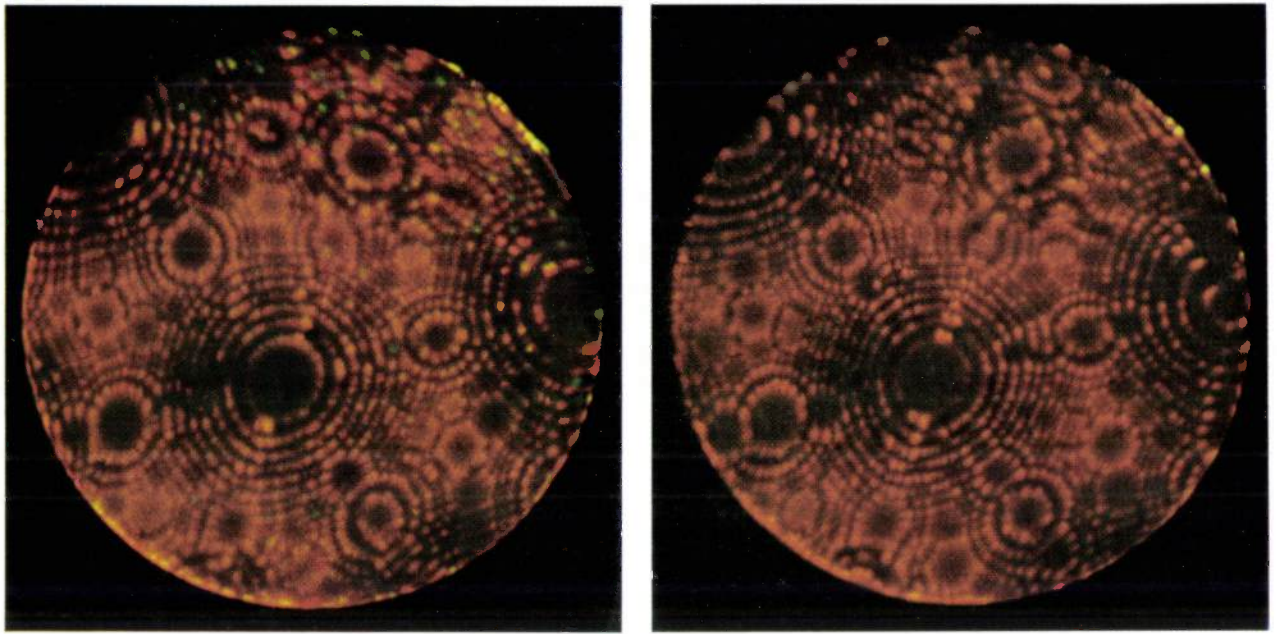
a   *b*

Fig. 18. Micrograph of a tungsten surface with 10^{-2} Pa neon at a voltage of 4600 V and a temperature of 78 K. 2.1×10^{-5} Pa nitrogen is also present. *a*) The field-strength has been lowered to zero for 10 seconds; *b*) the field-strength has been lowered to 23 GV/m for the same period. Both micrographs show the surface in red before the field-strength reduction and in green after the field-strength reduction. In (*a*) adsorption and hence reorientation were apparently possible, as can be seen from the red and green atoms which are visible (right and top). This was not the case in (*b*) where only yellow atoms can be seen.

If τ is the lifetime of an atom in the ionization zone, then:

$$\tau = 1/D\nu, \tag{4}$$

where D is the tunnelling probability as given by eq. (2) and ν the Bohr frequency of the electron in the atom. The ionization probability $P(t)$ during the time interval t can be written as

$$P(t) = 1 - \exp\left(-\frac{t}{\tau}\right). \tag{5}$$

The ionization probability in the entire space outside a shell with a thickness x_c (see fig. 11) around the specimen is then given by:

$$P(t) = 1 - \exp\left[-\int_{x_c}^{\infty} \frac{\nu D(x)}{v(x)} \, dx\right], \tag{6}$$

where $v(x)$ represents the velocity of the gas atom.

The number of molecules Z_s reaching the ionization zone from the gaseous phase under the influence of the electric field is given by eq. (3). The total number of molecules Z that can penetrate to the surface is given by:

$$Z = Z_s \exp\left[-\int_{x_c}^{\infty} \frac{\nu D(x)}{v(x)} \, dx\right]. \tag{7}$$

At low field-strengths $D(x)$ is zero over the entire inte-

gration area and therefore $Z = Z_s$. If the field-strength is increased, field ionization will occur and $D(x)$ will increase rapidly. This will cause Z to decrease, finally becoming zero at complete field ionization.

This effect of the field-strength on field ionization can also be found experimentally. *Fig. 19* is a graph of the field-ion current i as a function of the electric field-strength E for a nitrogen pressure of 4.5×10^{-2} Pa measured for the tungsten specimen shown in fig. 17. The field-strength was calibrated by establishing the value required for field evaporation at 10^{-2} Pa helium at 78 K, which is known to be 55 GV/m. The ionization probability P_i is also plotted in the figure.

The curve in fig. 19 shows the same behaviour as the curve for the field ionization of inert gases. In the initial part of the curve $d(\log i)/d(\log U) = 36$; similar values were found for helium, neon and argon. It was decided to use a linear horizontal axis in the figure to enable the field-strengths at which field ionization occurs to be plotted more accurately.

Comparison with the theoretical values of the field-strength for the nitrogen atom and the nitrogen molecule, as given in Table II, shows that the micrograph in fig. 17 was obtained with the aid of nitrogen *molecules*.

Fig. 19 will make it easier to appreciate why in the experiment illustrated in fig. 18, in which the electric field was reduced to 23 GV/m, no adsorption occurred.

All the nitrogen molecules have been ionized at this field-strength. The same applies to the experiment illustrated in fig. 17, in which the specimen was imaged in nitrogen at a field-strength of 18.5 GV/m.

It is possible to lower the field-strength periodically briefly by means of high-voltage pulses without any noticeable effect on the image. In one experiment we used rectangular pulses of 1 μs width and a repetition
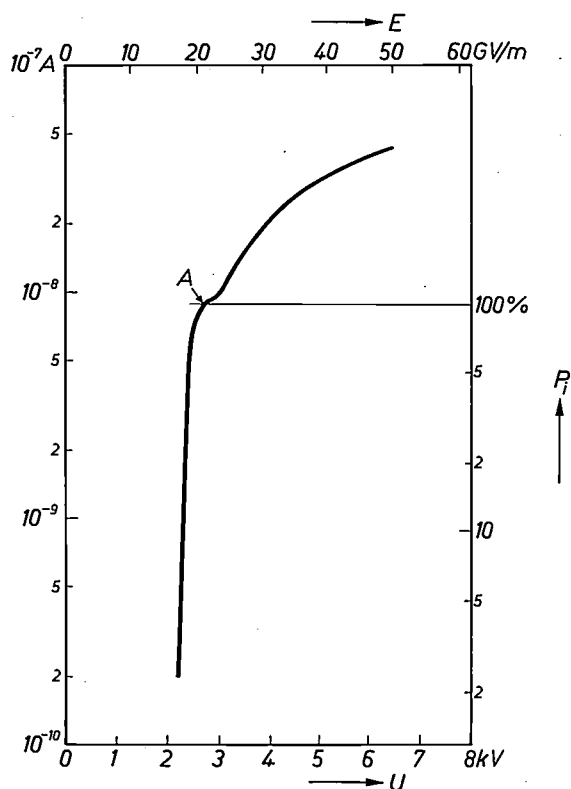
Fig. 19. Field-ion current i as a function of the voltage U (and of the field-strength E) for a nitrogen pressure of 4×10^{-2} Pa measured on the tungsten specimen shown in fig. 17. If it is assumed that all the molecules have been ionized at the knee A of the curve, the ionization probability P_i can be plotted vertically. The calculated ionization probability is in good agreement with the curve for this field-strength and for the ionization energy of nitrogen.

cules per cm² were adsorbed during the same period. During imaging with neon this gas was adsorbed on the surface by polarization forces. It was desorbed again during the pulse because the bond energy was then too small. At 78 K neon is bound only at field-strengths between 23 and 50 GV/m. Adsorbed neon cannot, therefore, have any effect on the nitrogen adsorption process, which takes place at lower field-strengths. Under the influence of the electric field the adsorbed nitrogen atoms will cause reorientation and thus reveal their presence.

In the experiment the field-strength was varied between 12 and 24 GV/m during the pulse. The same part of the surface, with an area of 8×10^{-13} cm², was considered for each 15 seconds of adsorption. The number of new spots occurring was recorded with the colour technique. The number N_m thus found is shown in *Table III* as a function of the field-strength E during the pulse. The table also includes the function Z_s

Table III. Summary of the data from an experiment described in the text, in which the effect of the field-strength on the ionization probability of the image gas in an FIM was investigated by briefly lowering the field-strength at periodic intervals. N_m is the number of observed new spots occurring as a result of non-ionized molecules being adsorbed. Z_s is the expected number of new spots calculated with eq. (3), on the assumption that no molecule is ionized, that all N_2 molecules dissociate after adsorption and that every atom causes a new spot. The number registered N_m is smaller for higher values of the field-strength E during the pulse because a proportion of the molecules are prematurely ionized (probability P_i) and therefore fail to reach the surface.

E (GV/m)	Z_s	N_m	P_i
0	3	3	0
12.7	21	19	0
14.6	24	14	0
15.6	25	21	0
17.6	29	4	7
19.0	31	2	90
21.4	35	1	100
23.3	38	0	100

frequency of 10³. In addition to 10⁻² Pa neon the system included nitrogen with a pressure of 2.1×10^{-5} Pa. The amplitude of the pulse was variable and the field-strength could be set to any desired value between zero and 34 GV/m. If the pulse causes the field-strength in such a system to drop below the minimum value required for field ionization of nitrogen, nitrogen can be adsorbed while the pulse lasts. Because of the short pulse duration and low repetition frequency the adsorption rate in our experiment was low. In a period of 15 seconds only 2×10^{12} nitrogen molecules per cm² were adsorbed, while the adsorption rate for impurities was so low that not more than 10⁹ mole-

calculated with eq. (3), on the assumption that all the molecules dissociate and each atom becomes visible after the increase of field-strength. 'Visible', does not of course mean that the nitrogen atom can actually be seen. As already stated, it is probably only tungsten atoms that are seen. Also included in the table is the ionization probability P_i as a function of the field-strength.

The agreement between the calculated and measured values can be considered surprisingly good, particularly if the statistical fluctuations are borne in mind. Another interesting point is that at 19 GV/m only two of the possible 31 events were observed. If

this is compared with the field-ion current measured at this field-strength and the number of nitrogen molecules expected in the ionization zone, it is found that nearly 90% of the molecules are ionized, which is in reasonable agreement with the adsorption experiment.

The conclusion must therefore be that, even if only indirectly in the form of reorientation of the surface, the adsorption of a number of molecules can be observed and this number is in good agreement with the number calculated.

Summary. Field-electron emission is the effect in which a metal surface exposed to sufficiently strong electrical fields emits electrons. To obtain emission at voltages which can be conveniently used in practice, the cathode has to be a tip with a sufficiently small radius of curvature. In the field-emission microscope the specimen acts as the cathode and the emitted electrons image the specimen tip in stereographic projection on a fluorescent screen located opposite it. To prevent adsorption at the surface an ultra-high vacuum is necessary. Field emission of ions can be obtained in a similar system if the direction of the field is reversed, the field-strength increased and a small quantity of gas, e.g. helium, admitted to the vacuum chamber. On arrival in the vicinity of the specimen surface the gas molecules will each give up an electron to the surface and the resultant ions will then leave the surface. Images of the specimen tip are thus obtained in which the atoms at the surface are individually visible. The use of a channel plate as an image intensifier has made it possible to increase the brightness of the screen and also to use other image gases. The channel plate also renders it possible to apply special colour techniques. The article shows that both methods of field emission are suitable for qualitative as well as quantitative investigation. The work done on field electron emission has also contributed to understanding of the electrical discharge between electrodes.

Magnetic liquids

The liquid shown in the photographs is behaving in the most curious manner: it is rising out of the container towards the rod above. It can be seen that some of the liquid has already passed from the container to the rod. Under certain conditions, a stable bridge of liquid can be formed between the rod and the liquid in the container, as the third photograph shows. The rod is a permanent magnet and the black liquid is a magnetic liquid, a colloidal suspension of very small particles of magnetite, Fe_3O_4. Owing to their minute dimensions (7-10 nm), these particles are superparamagnetic, i.e. they have a saturation magnetization just like a ferromagnetic material but their direction of magnetization fluctuates continuously under the influence of the thermal agitation. In addition, the particles undergo the usual Brownian movement of small particles in a fluid. However, when the particles are subjected to a magnetic field, they all assume practically the same direction of magnetization. The liquid therefore behaves magnetically as soft iron.

To prevent flocculation of the particles they must be coated with some surface-active substance. After this treatment the excess surface-active agent must be removed. At Philips Research Laboratories, J. M. A. Nevelsteen and Dr P. C. Scholten have devised an improved method for this, which will give a very high concentration of particles in the magnetic liquid. In this way magnetic liquids with a saturation magnetization of up to 0.06 T can be prepared.

Apart from the curious behaviour shown in the photographs, magnetic liquids possess other interesting properties. They can for example be set in motion by a moving magnetic field, and their viscosity is field dependent.

Most applications of magnetic liquids depend on one of the following: the ability of a magnet to hold the liquid in some fixed configuration; influencing the streaming of a liquid; and the possibility of injecting magnetic signals into a liquid. Examples of the first possibility are rotating seals and new types of fluid bearing.
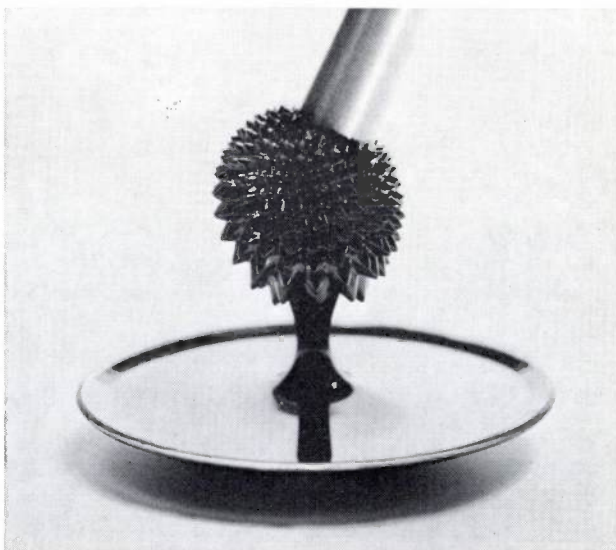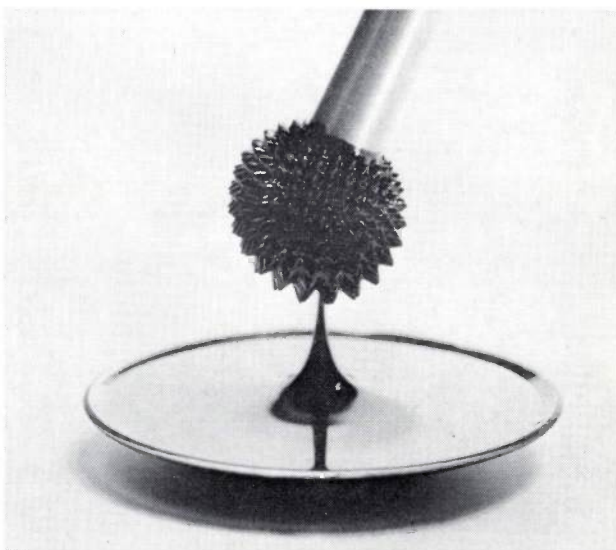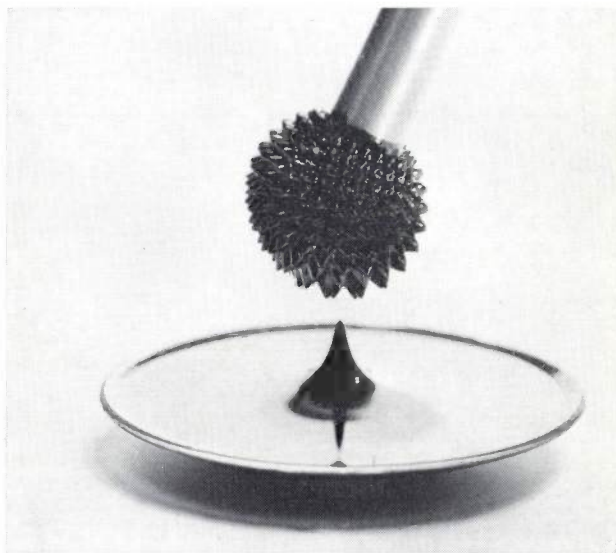
Some inductorless filters

D. Blom, H. W. Hanneman and J. O. Voorman

The frequency-dependent elements that have always been used in filters are capacitors and inductors. With the current trends for miniaturization of circuits, it has become increasingly difficult to use inductors, first in low-frequency circuits and later also at higher frequencies. The need for networks with no inductors soon became apparent; these networks have to include circuits that simulate the inductors. One way of doing this is by introducing active elements. This article describes three types of these 'active' filters.

Introduction

Filters are used in practically all electronic systems; from the very beginning of the science now called electronics they have occupied a central place. In the very first volume of this journal a series of articles were published on this subject [1]. The classical filters discussed there were composed of *passive* elements: resistors, capacitors and inductors. The use of inductors is not strictly necessary: filters can be made from resistors and capacitors only (*RC* filters). However, when inductors are included it is easier to obtain some of the characteristics desired of a filter, such as high discrimination between desired and undesired signals. If such a filter were to be constructed entirely from resistors and capacitors, a large number of components would be required and the signal would be considerably attenuated. On the other hand, inductors have a number of serious disadvantages: they are usually larger, heavier and more expensive than capacitors of the same impedance and Q (quality factor). This is especially the case for inductors for low frequencies, which often need a ferrite core. Since these filters cannot be built into integrated circuits — because resistance values in integrated circuits cannot be sufficiently accurately defined and because suitable inductors and capacitors cannot be made with values in the range of interest — they have to be combined with ICs as separate elements. The objection to inductors has therefore increased steadily over the years: their use makes the size of the filter disproportionately large compared with the rest of the circuit.

For certain applications — in fact inductors can sometimes be used with advantage — a need has therefore arisen for filters without inductors, which nevertheless give the same characteristic behaviour as filters

with inductors. One way in which this may be achieved is to introduce one or more *active* elements, i.e. amplifiers, into the circuit. Some filters of this type were described about ten years ago in this journal by G. Klein and J. J. Zaalberg van Zelst [2].

In recent years the techniques used in integrated circuits have been used to obtain a better solution to the problem; with these techniques very compact circuits can be constructed that contain active elements and can replace the inductors.

From recent publications on active *RC* filters [3] it is evident that the design of such a filter can be based on a variety of quite different circuits. In this article we shall describe three types of circuit investigated at Philips Research Laboratories. First, however we shall give a review of general ideas on filters, especially those that have played a role in the choice of the circuits described here.

The transfer function of a two-port network

Most filters are networks with two pairs of terminals (two-ports); one pair is considered as the input port and the other as the output port (*fig. 1*). The ratio of the complex expressions for the input and the output quantities (e.g. the voltages) is termed the *transfer function*. If the circuit is entirely composed of elements whose dimensions are negligible in comparison with the wavelength of the signals, the transfer function can
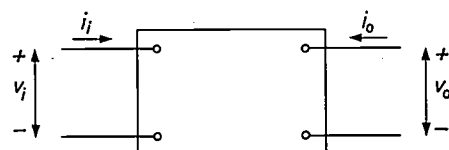
Fig. 1. Two-port circuit. v_i input voltage. v_o output voltage. i_i input current. i_o output current. These quantities are taken as positive in the directions indicated.

Ir D. Blom, Ir H. W. Hanneman and Ir J. O. Voorman are with Philips Research Laboratories, Eindhoven.

be expressed as the ratio of two polynomials [4]:

$$\frac{V_0}{V_1} = \frac{b_m p^m + b_{m-1} p^{m-1} + \ldots + b_1 p + b_0}{a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0}. \quad (1)$$

In this expression p, sometimes called the Laplace operator, is equal for sinusoidal signals to the product of $j = \sqrt{-1}$ and the angular frequency:

$$p = j\omega. \quad (2)$$

Further, $a_0 \ldots a_n$ and $b_0 \ldots b_m$ are real constants. The transfer function is thus a real rational function of p. The highest power of p occurring in the expression is called the order of the function and of the corresponding filter.

Expression (1) can also be written in the form:

$$\frac{V_0}{V_1} = A \frac{(p - z_1)(p - z_2) \ldots (p - z_m)}{(p - p_1)(p - p_2) \ldots (p - p_n)}, \quad (3)$$

where A is a constant; $z_1 \ldots z_m$ are the zeros and $p_1 \ldots p_n$ the poles of the function. Because the coefficients in the polynomials are real, the poles and zeros that are not real occur in complex-conjugate pairs.

If the number of poles is greater than the number of zeros ($n > m$), the transfer function tends to zero for high frequencies. The corresponding network then forms a *lowpass filter*. We shall be chiefly concerned with this type in this article, since expressions and sometimes circuits, for highpass filters, bandpass filters or bandstop filters can be derived in a simple way from those for lowpass filters.

The simplest transfer function for a lowpass filter is one in which no zeros occur at finite frequencies; p does not then appear in the numerator. A filter of order n will in this case be characterized by:

$$\frac{V_0}{V_1} = \frac{1}{a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0}. \quad (4)$$

The characteristics of an ideal lowpass filter

The purpose for which a filter is intended determines the requirements demanded of it. If it is intended merely to discriminate between signals of different frequencies, then only the *amplitude characteristic* is of interest; this is the ratio of the amplitude of the output signal to that of the input signal (i.e. the modulus of (4)) as a function of frequency. The phase shift that takes place in the filter is then of secondary importance. Usually it is required that the transmission loss or gain of the filter should be the same for all frequencies in the passband, i.e. that the amplitude characteristic in the passband should be as flat as possible. In the stopband a certain minimum attenuation is required. An ideal amplitude characteristic therefore has the form shown in *fig. 2a*. The quantity Ω plotted hori-

zontally is the frequency ω normalized with respect to the cut-off frequency ω_0.

In other cases, however, the specifications relating to the phase shift may be more important. Usually the phase shift is required to be as accurately proportional to the frequency as possible. The time delay of the signal (the derivative of the phase shift with respect to the frequency) is then independent of the frequency. Obviously this is only of importance in the passband of the filter. The ideal time *delay characteristic* is thus a horizontal straight line in the passband (fig. 2b).
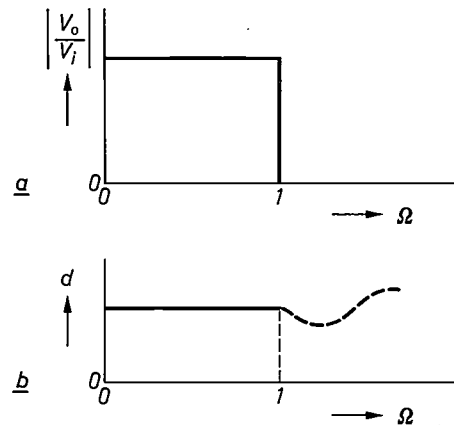
Fig. 2. *a*) Ideal amplitude characteristic of a lowpass filter. The modulus of the transfer function V_0/V_1 is constant in the transmission region. Ω is the normalized frequency, i.e. the frequency ω divided by the cut-off frequency ω_0. *b*) Ideal time-delay characteristic for a lowpass filter. The time delay d is constant in the passband.

Common types of transfer characteristic

A filter with an ideal amplitude or delay characteristic would have to have an infinite number of elements. In practice therefore a compromise must be made. We shall now look at four types of characteristic that are often used.

The simplest types are the *maximally flat or Butterworth characteristics*. These are amplitude characteristics whose shape is such that for a filter of order n, at zero frequency, the 1st, 2nd, ... nth derivatives with respect to frequency are zero. *Fig. 3a* shows the shape of the amplitude characteristics for various values of n. The corresponding time-delay characteristics are shown in fig. 3b.

[1] B. van der Pol and T. J. Weijers, Electrical filters, Philips tech. Rev. 1, 240-245, 270-276, 298-306, 327-334 and 363-366, 1936.

[2] G. Klein and J. J. Zaalberg van Zelst, Some simple active filters for low frequencies, Philips tech. Rev. 25, 330-340, 1963/64.

[3] A recent and fairly full bibliography has been given by S. K. Mitra, in: Analysis and synthesis of linear active networks, Wiley, New York 1969.

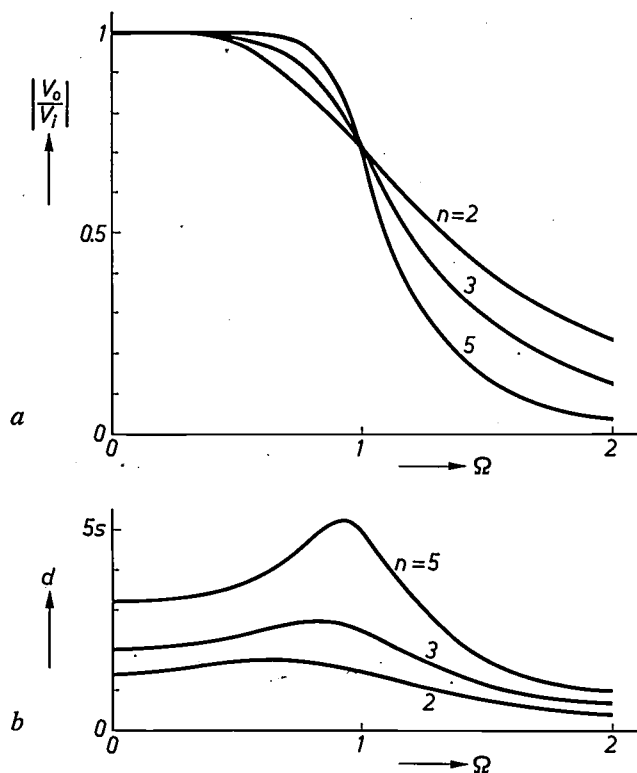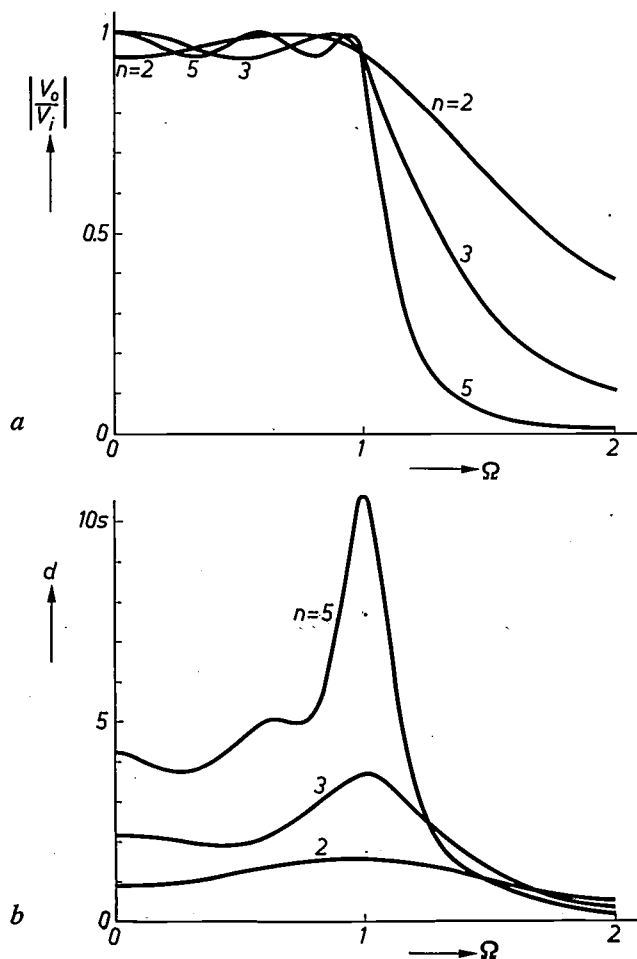[4] See for example E. A. Guillemin, Synthesis of passive networks, Wiley, New York 1959 (2nd printing).

Fig. 3. a) Amplitude characteristics of Butterworth filters for three values of the order n. b) Time-delay characteristics of these filters.

Butterworth functions have the form given by (4) and for sinusoidal signals the modulus of a Butterworth function is given by:

$$\left|\frac{V_o}{V_i}\right| = \frac{1}{\sqrt{(1 + \Omega^{2n})}}, \tag{5}$$

where Ω is again the normalized frequency.

The values of the coefficients required in (4) to enable the modulus to satisfy (5) have been published in tables [5].

A Butterworth curve gives the best approximation to the ideal characteristic at $\Omega = 0$, but at frequencies near the cut-off frequency it is not so good an approximation.

An approximation that is equally good over the whole passband can be achieved with *Chebyshev characteristics*. These characteristics do not fall away monotonically with increasing frequency but have a ripple in the passband. The number of peaks is equal to $n/2$. The peaks are all of equal height, and so are the troughs. Such a curve is therefore called an equal-ripple characteristic. *Fig. 4a* shows such characteristics for various values of n. In this example the depth of the ripples is $\frac{1}{2}$ dB. If the ripples in the curve are made deeper, there is a steeper cut-off. The time-delay characteristics are given in fig. 4b.

Chebyshev characteristics can also be represented by (4). The values that the coefficients then have to assume are again given in tables [5].

When the time delay is required to remain substantially constant as the frequency changes (as in the processing of pulse signals and in television circuits), Butterworth and Chebyshev filters are not satisfactory, as can be seen from figs. 3b and 4b. In such cases Bessel functions are frequently used in the design of filters. Filters of this type are called *Bessel filters*. Their particular feature is that, for a filter of order n, the time-delay characteristic has such a form that at zero frequency the 1st ... nth derivatives are all zero. (So that the time-delay characteristic is then maximally flat.) The transfer characteristic of Bessel filters can again be represented by the relation (4) and again there are tables giving the values of the coefficients [5]. *Fig. 5a* gives the amplitude characteristics and fig. 5b the phase characteristics of Bessel filters for various values of n.

The signal-separating performance of a filter can be improved by using a circuit in which the transfer function has not only poles in the passband but also zeros

[5]   See for example L. Weinberg, Network analysis and synthesis, McGraw-Hill, New York 1962.

Fig. 4. a) Amplitude characteristics of Chebyshev filters for three values of n. In this example, the ripple in the passband is $\frac{1}{2}$ dB. b) Time-delay characteristics.
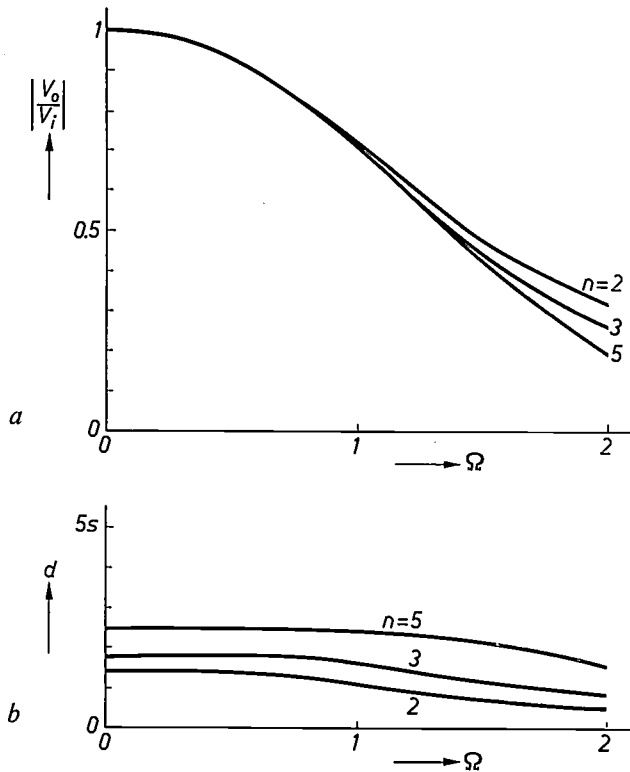
Fig. 5. *a)* Amplitude characteristics of Bessel filters for three values of *n. b)* Time-delay characteristics.
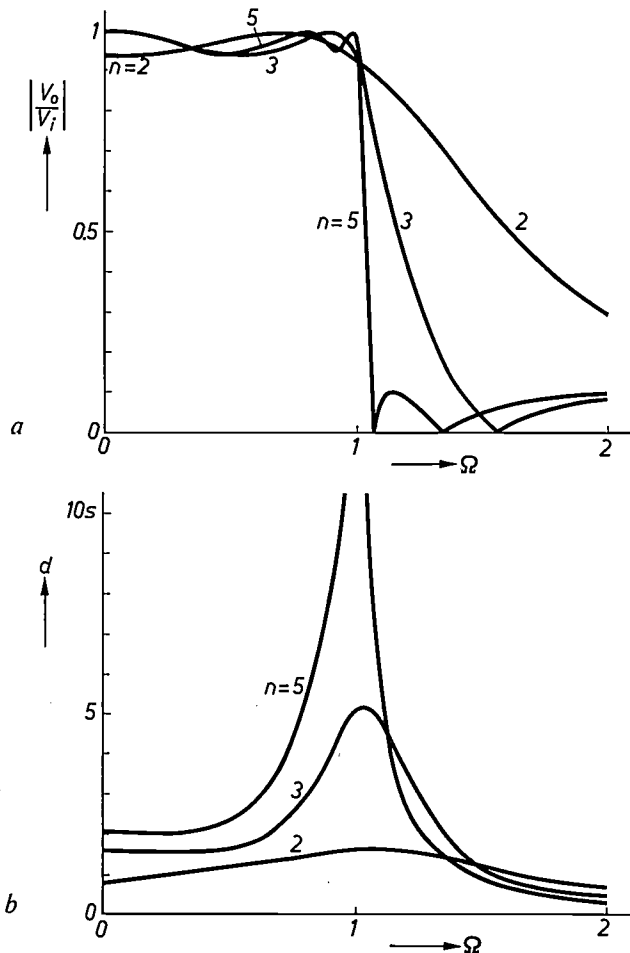
in the stopband. If these zeros occur at certain purely imaginary values of *p*, then the circuit is called an *elliptic filter* or a *Cauer filter*. A zero in the stopband, not too far from the cut-off frequency, makes the cut-off much steeper. This can be seen from *fig. 6a* which shows the amplitude characteristics of an elliptic filter for various values of *n*. The curves show ripples of equal height both in the passband and in the stopband (equal-ripple in both bands). The corresponding time-delay characteristics are shown in fig. 6*b*.

Passive filters

The simplest kind of lowpass filter consists of a resistor and a capacitor (*fig. 7*). The transfer function is:

$$\frac{V_o}{V_i} = \frac{1}{1 + p\tau}, \tag{6}$$
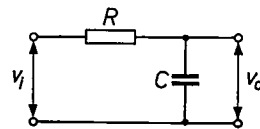
Fig. 7. Lowpass filter consisting of one *RC* section.

where $\tau = RC$. This is a Butterworth function of the first order with a cut-off frequency of $1/\tau$. A Butterworth function of higher order cannot however be obtained with a circuit containing only resistors and capacitors. For example, the transfer function of a circuit as shown in *fig. 8* is:

$$\frac{V_o}{V_i} = \frac{1}{1 + (\tau_1 + \tau_2 + \tau_2 C_1/C_2)p + \tau_1\tau_2 p^2}, \tag{7}$$
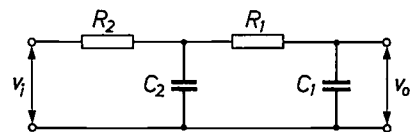
Fig. 8. Filter formed from two *RC* sections. A Butterworth characteristic cannot be obtained with this type of filter.

which is not a Butterworth function. This can be seen from the fact that the function (7) only has real poles, whereas in a Butterworth function of order 2 the poles form a complex conjugate pair.

Fig. 6. *a)* Amplitude characteristics of elliptic filters for three values of *n*. In both the passband and stopband there are ripples of constant depth. In this example the ripple in the passband is $\frac{1}{2}$ dB; in the stopband the ripple depth is sufficiently great for the suppression of the signals to be at least 20 dB compared with the passband. *b)* Time-delay characteristics.

Circuits with one or more inductors give much more scope. For example, the transfer function of the circuit of *fig. 9* is:

$$\frac{V_o}{V_i} = \frac{1}{1 + p\tau + p^2/\omega_r^2},\qquad(8)$$
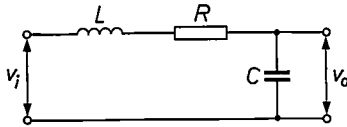
Fig. 9. Filter section consisting of inductance, resistance and capacitance. Butterworth, Chebyshev and Bessel characteristics can be obtained with this circuit.

where $\omega_r^2 = 1/LC$ and $\tau = RC$. Putting $\tau = \sqrt{2}$ and $\omega_r^2 = 1$ gives a Butterworth filter of order 2 with a cut-off frequency of 1 radian/s. A different cut-off frequency is obtained by altering τ in inverse proportion to the desired cut-off frequency and altering ω_r in direct proportion. With a different choice for τ and ω_r Chebyshev or Bessel filters of order 2 can be obtained.

Active filters

By using one or more amplifiers, RC networks can be made that behave in the same way as networks with inductance. A simple example is shown in *fig. 10*. This circuit, proposed by R. P. Sallen and E. L. Key [6], contains an amplifier of voltage gain K. Assuming that this amplifier has an infinite input impedance and that its output impedance is zero, analysis of the circuit shows that the transfer function of the circuit is:

$$\frac{V_o}{V_i} = \frac{K}{1 + \left\{\tau_1 + \tau_2(1 + C_1/C_2 - K)\right\}p + \tau_1\tau_2p^2}.\qquad(9)$$

Choosing the elements correctly will give the transfer function of a Butterworth, Chebyshev or Bessel filter. If, for example, we make $R_1 = R_2 = 1\ \Omega$ and $C_1 = C_2 = 1$ F, then $\tau_1 = \tau_2 = 1$ s; with a voltage gain $K = 3 - \sqrt{2}$ we then have the transfer function:

$$\frac{V_o}{V_i} = \frac{1.586}{1 + p\sqrt{2} + p^2}.\qquad(10)$$
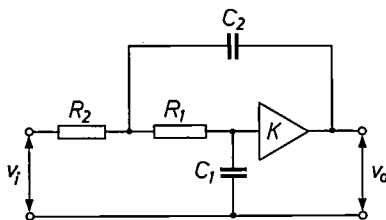
Fig. 10. Active filter of the second order after R. P. Sallen and E. L. Key [6]. This filter differs from classical passive filters by the presence of an amplifier.

This is a Butterworth function of order 2 with a cut-off frequency of 1 radian/s. Filters with a different cut-off frequency can again be obtained by changing τ_1 and τ_2 in inverse proportion to the desired cut-off frequency.

Criteria for the choice of a filter

In designing a filter, the basic requirements are those set by the amplitude and phase characteristics. For classical filters, tables and graphs have been published relating to the standard characteristics and for most of the corresponding circuits [7]. In this way computations can be reduced to a minimum. Similarly, in applying active RC filters it is preferable to take a circuit whose component values can be simply derived from existing tables. If this is not the case, for example because a non-standard transfer characteristic may be required, a circuit is of course chosen whose component values can be calculated easily.

Apart from these criteria there are other significant factors in the choice of circuit. Of great importance, for example, is the sensitivity of the characteristic to deviations of the component values from their desired values; thus, with a filter of low *parameter sensitivity* a wider tolerance on the component values is admissible. Also, with components of a given tolerance, a more accurate filter can be made. In classical filters these requirements can readily be satisfied, largely because they are usually ladder networks, which are less sensitive to small deviations in component values than two-ports of other configurations. The low loss of practical capacitors and inductors used in filters also helps to give low parameter sensitivity.

This low parameter sensitivity of low-loss filters can be understood by considering a lossless two-port circuit that connects a signal source of internal resistance R_i to a load resistance R_l [8] (*fig. 11*). If the signal source is optimally matched, i.e. the maximum power is dissipated in R_l, then any change in parameter values of the two-port circuit leads to a decrease in the power dissipated in R_l. The first derivative of the output signal amplitude with respect to the value of any circuit element will then be zero; this implies a minimum sensitivity of the signal transmission ratio to variations in the components.

Deviations of the parameters in a filter occur as a result of the normal spread of values of the components, ageing of the components, temperature changes and the effect of moisture. In active filters, deviations in
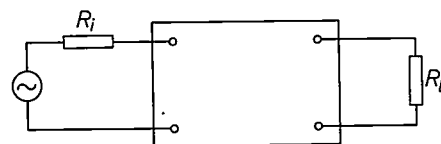
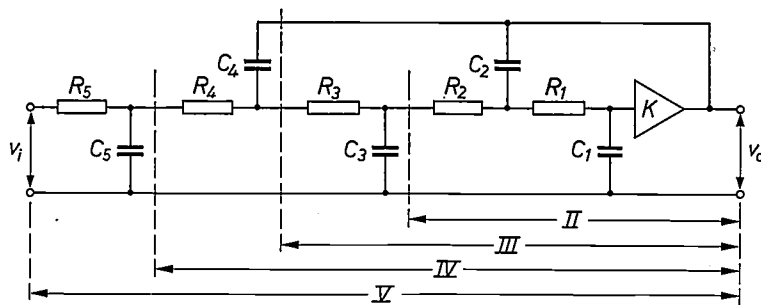Fig. 11. Two-port with load resistance R_l, connected to a voltage source of internal resistance R_i.

Fig. 12. Filters of the 2nd ... 5th orders (marked by II to V) based on the same principle as the filter of fig. 10.

amplifier parameters also have their effect; such deviations can arise because of changes in the supply voltage.

The sensitivity of filters to parameter changes can be of vital importance in the choice of a filter. A number of studies have consequently been devoted to this topic [9].

Another feature that may be significant in the choice of a filter is the signal-voltage range for distortion-free processing by the filter — the dynamic range. In filters composed of passive elements there are in principle no limits on the size of the signal [10]. In an active filter, however, the amplitude of the signal is restricted because the amplifiers must not be overloaded. In most filters the output voltage in the passband is approximately equal to the input voltage. This is however not the case for the voltages occurring *inside* the filter. In active filters of higher order, voltages much larger than the input voltage can occur. The input voltage must of course be limited to such a value that no single amplifier is overloaded. Because the filter introduces noise, this implies a limitation on the signal-to-noise ratio. This must also be taken into account in the design of active filters.

Finally there is one rather obvious criterion that can affect the choice of the type of filter; this is the simplicity of the circuitry. In an active filter it is preferable to keep the number of elements small.

These various features have played a role in the choice of the three types of circuit for the active *RC* filters investigated and discussed below. In the first type the simplicity of the circuit was the primary criterion; these filters, including those of higher order, contain only one amplifier. The second type discussed is characterized by a very simple design procedure; because little computation is required, transfer characteristics of non-standard type can be obtained without much difficulty. This is especially important for applications in the laboratory. Finally, we shall discuss a type of circuit which can be used to give very accurate filters. In these circuits gyrator elements are used to simulate inductors with very low losses. As we saw above, this allows the parameter sensitivity of the circuit to be kept to a very low value.

Active filters with one amplifier

A method for the design of higher-order filters using only one amplifier has been given by T. Fjällbrant [11]. *Fig. 12* shows a scheme that he has proposed for filters up to the fifth order. In his article he demonstrates a method of calculating the filter elements required to obtain a given transfer function. Fjällbrant only considers the case in which the input and output voltages of the amplifier are equal ($K = 1$). There are certain disadvantages to this method: the calculations are rather lengthy and all the resistances and capacitances are given different values. Another more practical disadvantage is that very large and very small resistances and capacitances are required.

The difficulties with Fjällbrant's approach largely disappear if the requirement of an amplifier with a voltage gain of unity is abandoned [12]. In this case, for a circuit of order n, we have $2n + 1$ variables ($R_1 \ldots R_n$, $C_1 \ldots C_n$ and the gain K). Since, to obtain a transfer function of order n, we have only n conditions to be fulfilled, $n + 1$ quantities can be chosen at will. For example, we can make all the resistances $1\ \Omega$ and two capacitances can be made equal to one another (for a filter of order n, $C_n = C_{n-1}$). For reasons that will be given later filters of this type are usually restricted to an order of five. The transfer function of a filter of order five is:

[6] R. P. Sallen and E. L. Key, A practical method of designing *RC* active filters, IRE Trans. CT-2, 74-85, 1955.

[7] See for example P. R. Geffe, Simplified modern filter design, Iliffe Student Edition, London 1964.

[8] See H. J. Orchard, Inductorless filters, Electronics Letters 2, 224-225, 1966.

[9] See for example P. J. McVey, Sensitivity in some simple *RC* active networks, Proc. IEE 112, 1263-1269, 1965;
H. W. Hanneman, The systematic and the random errors due to element tolerances of electrical networks, Philips Res. Repts. 26, 414-423, 1971;
H. W. Hanneman and H. N. Linssen, A relation between sensitivity and variance of network functions, IEEE Trans. CT-19, 499-502, 1972.

[10] In telephony filters, in which very small inductors are often used, the signal level may however be limited by distortion arising from saturation of the core.

[11] T. Fjällbrant, Canonical active *RC* filters with a low transfer function sensitivity, Ericsson Technics 23, 211-238, 1967.

[12] See H. W. Hanneman, Higher-order *RC*-active filters, Philips Res. Repts. 26, 65-74, 1971.

$$\frac{V_0}{V_1} = \frac{K}{a_5 p^5 + a_4 p^4 + a_3 p^3 + a_2 p^2 + a_1 p + 1} . \quad (11)$$

Depending on whether a Butterworth, Chebyshev or Bessel function is required, the values of the coefficients $a_1 \ldots a_5$ are selected from the appropriate tables [5]. Analysis of the circuit then permits a set of five non-linear inhomogeneous equations to be written for the five variables $C_1 \ldots C_4$ $(= C_5)$ and K. The values of the capacitances and the gains for the three types of filter to obtain a cut-off frequency $\omega_0 = 1$ radian/s are given in *Table I*. It can be seen that the capacitances no longer differ greatly in value.

The same cut-off frequency is obtained if all the resistances are multiplied by the same factor and all the capacitances are multiplied by the reciprocal of this factor. For an arbitrary cut-off frequency ω_0, if all the resistances are kept constant, all capacitances must be changed in inverse proportion to ω_0 (or, if the capacitances are kept constant, the resistances must be changed).

The gain required has a value between 1 and 2. This can be provided by an operational amplifier with voltage feedback from the output to one of the input terminals (*fig. 13a*). The gain is then almost completely determined by the feedback ratio, here equal to $(R_1 + R_2)/R_1$. Another circuit is shown in fig. 13*b*. In this case the gain can be adjusted by means of the variable resistance.

Highpass filters

Highpass RC filters can be derived quite simply from lowpass filters. It is only necessary to interchange the capacitances and resistances. In this way circuits for highpass filters of order 2 . . . 5 take on the form shown in *fig. 14*. The relation for the transfer characteristic can be derived from (4) by putting $1/p$ in place of p:

$$\frac{V_0}{V_1} = \frac{p^n}{a_n + a_{n-1} p + a_{n-2} p^2 + \ldots + a_1 p^{n-1} + a_0 p^n} \quad (12)$$

If all the capacitances have a value of 1 F and the gain has the value given in Table I, highpass filters with a cut-off frequency of 1 radian/s are obtained if the resistances have values equal to the reciprocals of the corresponding capacitances in Table I.

The parameter sensitivity of a large number of filters of the type now under discussion has been investigated. From these (computer) calculations it has been found that the parameter sensitivity is always largest at or near the cut-off frequency. To judge the usefulness of a filter in this respect it is therefore usually sufficient to investigate the behaviour at the cut-off frequency. It has also been found that the parameter

Table I. Values of the capacitances C_{1-5} and the gain K required for a cut-off frequency $\omega_0 = 1$ radian/s in Butterworth, Chebyshev and Bessel filters of order n, as in fig. 12.

	n	C_1	C_2	C_3	$C_4 = C_5$	K
Butterworth	2	1.000	1.000			1.586
	3	0.426	1.531	1.531		1.269
	4	0.151	2.815	1.535	1.535	1.106
	5	0.171	1.462	1.251	1.790	1.335
Chebyshev ($\frac{1}{2}$ dB ripple)	2	0.812	0.812			1.841
	3	0.342	2.080	2.080		1.219
	4	0.255	1.782	2.409	2.409	1.404
	5	0.140	1.909	1.529	3.694	1.381
Chebyshev (1 dB ripple)	2	0.952	0.952			1.953
	3	0.321	2.524	2.524		1.192
	4	0.295	1.665	2.717	2.717	1.508
	5	0.127	2.159	1.532	4.397	1.281
Bessel	2	0.785	0.785			
	3	0.392	1.043	1.043		
	4	0.072	2.944	0.967	0.967	1.009
	5	0.097	1.118	0.736	1.065	1.203
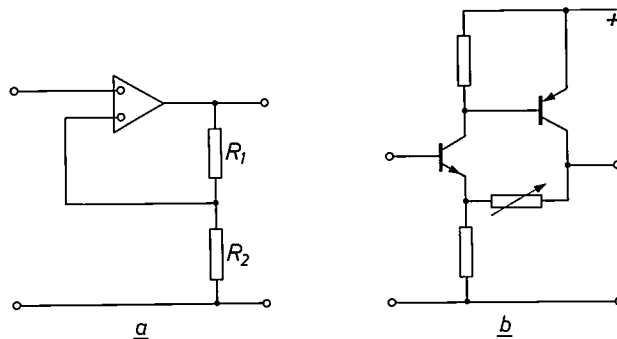
Fig. 13. Amplifiers for use in the active RC filters of figs. 10 and 12. *a*) Operational amplifier with negative feedback. *b*) Amplifier with adjustable gain.
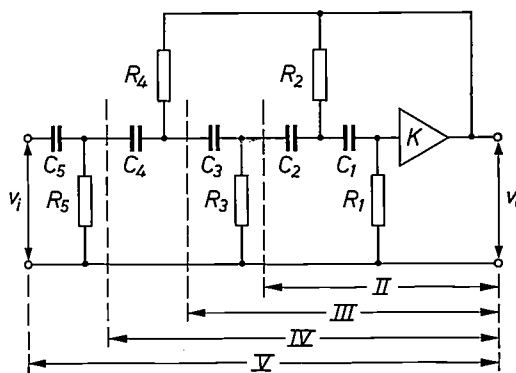

Fig. 14. Highpass active RC filters of the second to the fifth orders.

sensitivity increases with the order of the filter. To quote some figures: in a Butterworth filter of order three, a variation of 1% in the most sensitive element gives a variation in the voltage transfer ratio of 1.8% at the cut-off frequency; if the filter is of order 5, this voltage variation is 6%. A Chebyshev filter ($\frac{1}{2}$ dB

ripple) is much more sensitive in this respect: the same variation in the most sensitive element gives a variation of 2.3 % in a third-order filter. This figure rises rapidly with the order of the filter; for the fifth order it is 30 %. Deviations in the values of the passive elements can be partially compensated by adjustment of the gain but the sensitivity to changes in the elements remains.

An impression of the changes that can occur in the amplitude characteristic as a result of the spread in values of the elements is given in *fig. 15*. The figure shows the characteristic for a Chebyshev filter of order five. The bars indicate the limits within which 66 % of the transfer characteristics lie if the values of all elements have a (uniformly distributed) spread of 1 %. The dashed line indicates the mean of the range of spread.

The fact that the parameter sensitivity of this type of filter rises rapidly with the order of the filter is one of the reasons why this method of designing is not used for filters of order higher than five. Another reason is that the voltage magnification in a filter increases with higher order. A further difficulty is that, for filters of

tions whose poles have a rather high Q [13]. This difficulty becomes less serious if the various sections are connected in a particular sequence. The section with the largest Q is connected after the section with the lowest Q. Then the sections whose poles have the next-to-largest Q are connected after those with the next-to-lowest Q, and so on. A further considerable
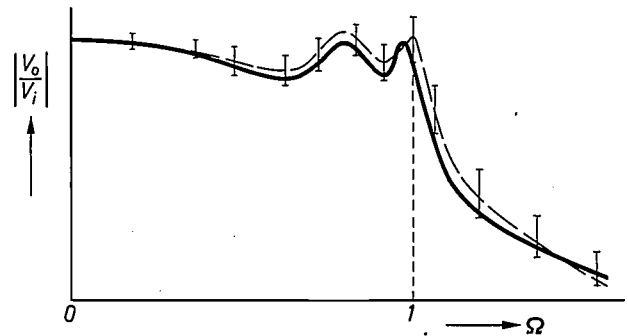
Fig. 15. Amplitude characteristic of a Chebyshev filter of the fifth order with a passband ripple of 1 dB. The bars at certain frequencies indicate the range of values of the transfer function obtained in 66 % of all cases when all the circuit elements have a uniform spread of 1 %. The dashed line indicates the mean of the range of spread.
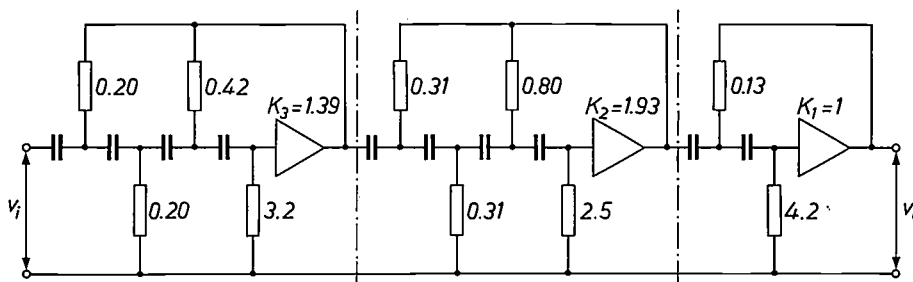
Fig. 16. Highpass filter of the tenth order consisting of two sections of fourth order and one section of second order (separated by chain-dotted lines). At the resistance values given the cut-off frequency is 1 radian/s; the capacitances then all have a value of 1 F.

higher order, very large or very small capacitances or resistances are necessary. We shall now discuss a method for satisfying the requirements for filters of higher orders.

Filters of higher order

A filter of higher order is obtained by cascading a number of sections of lower order. To design these sections the polynomial in the denominator of the transfer function of the filter is written as the product of a number of polynomials of lower order (this is always possible for rational functions).

In the article by Sallen and Key referred to earlier [6] it has been shown that a filter of the tenth order, for example, can be split up in this way into five sections of order two. A difficulty with this configuration is that rather high voltages can occur in various sections (voltage magnification). This is the case in those sec-

improvement is obtained by replacing each pair of sections chosen in this way by a single section of order 4. (If the order of the filter is not a multiple of 4, then of course this will only be possible to a limited extent.) *Fig. 16* shows the circuit of a highpass filter of order 10 that consists of two such sections of order 4 and one section of order 2. Analysis of this circuit has shown that the voltage magnification is only 2.7 whereas a cascade circuit of second-order sections in the scheme due to Sallen and Key has a voltage magnification of 30.

Elliptic filters

In filters of the type discussed here it is easy to arrange that zeros appear in the stopband and therefore to make elliptic filters. This is done by providing

[13] The Q of a pole is the imaginary part of the pole divided by twice the real part.

an extra path for the signal through the filter to the output. In *fig. 17* this is illustrated by a filter of the third order. The input terminal of the filter is connected to the input of the amplifier via R_0 and C_0. It can be shown that the signal at the amplifier input is then zero at the frequency ω_1 if R_0 and C_0 have the following values:

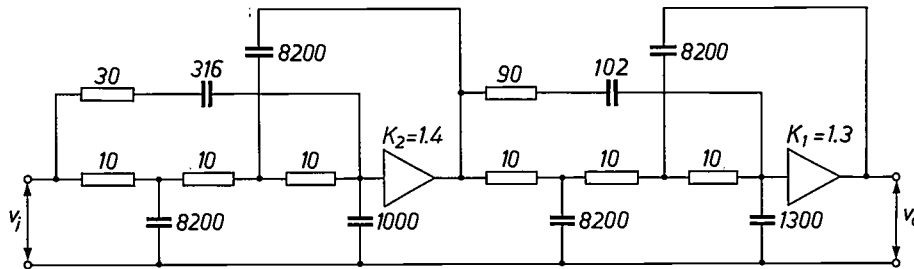$$R_0 = \omega_1{}^2 R_1 R_2 R_3 C_2 C_3 - R_1 - R_2 - R_3,$$

$$C_0 = \frac{1}{\omega_1{}^2 \left\{ R_1 C_2 (R_2 + R_3) + R_3 C_3 (R_1 + R_2) \right\}}.$$

The transfer function is then:

$$\frac{V_0}{V_i} = \frac{p^2 + \omega_1{}^2}{a_3 p^3 + a_2 p^2 + a_1 p + a_0}. \tag{13}$$

It can be seen from the numerator that there is indeed a zero at the frequency ω_1.

More than one zero can be obtained by combining a number of sections as in fig. 17. *Fig. 18* gives the circuit of such a filter consisting of two elliptic third-order sections. The cut-off frequency is 3.5 kHz and the zeros are at 5.0 and 6.6 kHz.

Fig. 17. Elliptical lowpass filter of the third order.

the output to one of the input terminals (*fig. 20*). This is sometimes called a *voltage follower*. Assuming, again, that the input impedance of the amplifiers is infinite and their output impedance zero, then analysis of the circuit yields the following transfer function:

$$\frac{V_0}{V_i} = \frac{1}{\tau_1 \tau_2 p^2 + \tau_1 p + 1}, \tag{14}$$

where $\tau_1 = R_1 C_1$ and $\tau_2 = R_2 C_2$. We obtain in this way a lowpass second-order filter of arbitrary type, provided τ_1 and τ_2 are chosen so that $\tau_1 = a_1$ and

Fig. 18. Lowpass filter consisting of two elliptic sections of the third order. The values of the resistances are in kΩ; capacitances are given in pF. The cut-off frequency is 3.5 kHz. The zeros in the stopband are at 5.0 and 6.6 kHz.

Filters to give an arbitrary transfer characteristic

In certain cases, for example in experimental work, the need arises for a filter with a transfer characteristic that does not correspond to one of the standard functions. Synthesis by the methods described above then usually involves much computational work because no tables exist for the values of the elements corresponding to the polynomials to be used. It is therefore desirable to have filters in which a simple relation exists between the circuit elements of the filter and the coefficients of the transfer function required.

We shall now describe a series of circuits which go a long way towards this. One way in which these circuits differ from those described above is that they contain more than one amplifier — which may be considered a disadvantage. The circuit of such a filter is shown in *fig. 19* [14]. Both amplifiers have a gain of unity. Operational amplifiers can be used with feedback from

Fig. 19. Lowpass active *RC* filter of the second order. Each amplifier has a gain of unity. With this configuration there is a very simple relation between the values of the circuit elements and the coefficients of the desired transfer function.
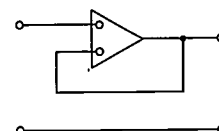
Fig. 20. Operational amplifier with negative feedback to give a gain very close to unity.

$\tau_1\tau_2 = a_2$, where a_1 and a_2 are the coefficients of the polynomial forming the denominator of the required transfer function.

Fig. 21 shows how a filter of arbitrary order can be designed on the principle of fig. 19. The transfer function is then given by (4), in which

$$a_1 = \tau_1, \ a_2 = \tau_1\tau_2, \ a_3 = \tau_1\tau_2\tau_3, \ldots a_n \rightleftharpoons \tau_1\tau_2 \ldots \tau_n. \tag{15}$$

For the design of a given filter we therefore have to choose the values of resistances and capacitances such that

$$\tau_1 = a_1, \ \tau_2 = a_2/a_1, \ \tau_3 = a_3/a_2, \ldots \tau_n = a_n/a_{n-1}. \tag{16}$$

There is then no objection to making all the resistances equal and the capacitance values can then be calculated by means of (16). Alternatively, the capacitances can all be made equal and the resistances then calculated.

In this way we of course find values for the time constants that again correspond to a cut-off frequency ω_0 of 1 radian/s. For a filter with a higher cut-off frequency, the time constants must be made smaller in inverse proportion to the required cut-off frequency.

voltage always occurs at or near the cut-off frequency. In practice, therefore, the investigation of voltage magnification is usually restricted to the cut-off frequency. An estimate of the voltages in the various amplifiers at the cut-off frequency can be made as follows: the voltage in the kth amplifier can be written:

$$V_k = \frac{a_{k-1}p^{k-1} + \ldots + a_1p_1 + a_0}{a_np^n + a_{n-1}p^{n-1} + \ldots + a_1p_1 + a_0} V_i. \tag{17}$$

Since the voltage transfer ratio is always approximately unity in the passband, the modulus of the denominator of (17) is also approximately unity. And since the normalized complex cut-off frequency is equal to j, the voltage magnification in the kth amplifier at this frequency is equal to:

$$\frac{V_k}{V_i} = j^{k-1}a_{k-1} + \ldots + a_4 - ja_3 - a_2 + ja_1 + a_0. \tag{18}$$

If $a_0 \ldots a_{k-1}$ are known, the function (18) can be depicted in a vector diagram that shows all the internal voltages in the filter. As an example, *fig. 22* gives these diagrams for Butterworth filters of the third and fifth orders. In these filters the largest voltage magnifications
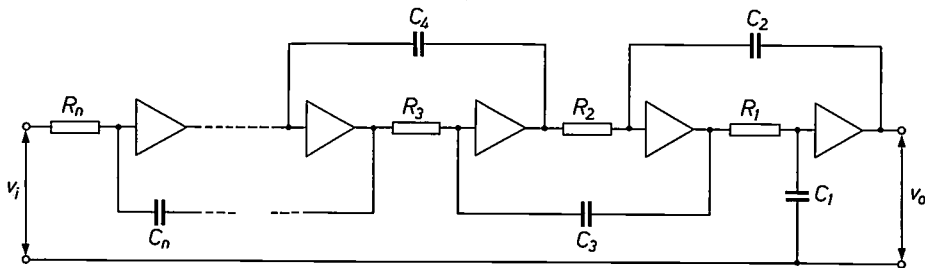
Fig. 21. Extension of the filter of fig. 19 to give a filter of arbitrary order.

Calculations show that the parameter sensitivity is again largest at or near the cut-off frequency. By changing the value of each element by 1%, we have investigated which of the elements has the greatest effect on the voltage transfer ratio at the cut-off frequency; the following figures relate to a 1% variation of the 'most-sensitive' element. In a Butterworth filter of the third order, the maximum change in the voltage transfer ratio is 1.5%; for a filter of the fifth order, it is 3.1%. A Chebyshev filter ($\frac{1}{2}$ dB ripple) of the third order shows a maximum variation in the voltage transfer ratio of 1.7%; for a fifth-order filter, it is 6.2%. These figures are better than those for the filters described earlier.

In filters of the type shown in fig. 21, every section contains an amplifier: the danger of overload is therefore greater than in the filters described earlier. Calculations have again shown that the maximum signal
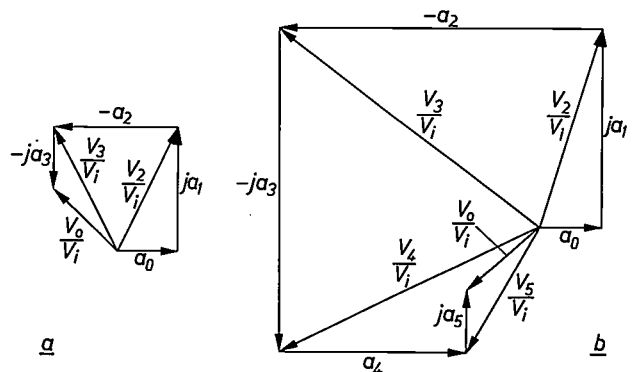

Fig. 22. Vectorial representation of the terms of (18), giving an impression of the voltage magnification at the cut-off frequency in the amplifiers of a filter of the type shown in fig. 21. *a*) Diagram for a Butterworth filter of the third order. *b*) Diagram for a Butterworth filter of order five.

[14] While this article was being prepared it was found that a similar circuit had been published by F. Anday: Active realization of nth-order low-pass transfer functions, Proc. IEEE 60, 909-910, 1972.

are respectively 1.6 and 3.7. For Chebyshev filters with a ripple of $\frac{1}{2}$ dB in the transmission region these figures are respectively 2.2 and 8.6. The figures are still larger for filters with a larger ripple. For example, in a Chebyshev filter of the fifth order and a ripple of 2 dB the voltage magnification has a value of 12.3.

The limitation imposed by the voltage magnification on the amplitude of the input signals and its unfavourable effect on the signal-to-noise ratio again often acts

section, the emitter of the other is connected to the capacitor of the previous section. If these transistors have a large enough current gain, the collector and emitter currents are effectively equal and the collector currents of the upper transistors will be equal to the currents in the resistors. In the circuit shown, the sum of these currents flows in the resistor R_t. If the output is taken from this resistor, then the transfer function is:
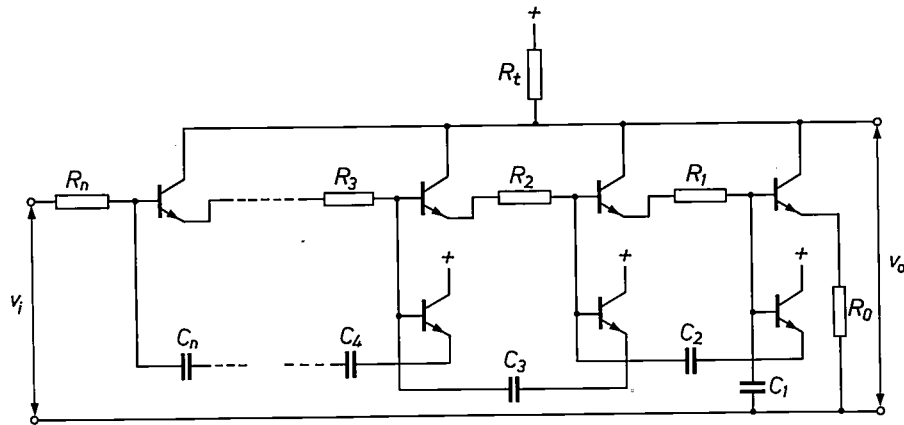
Fig. 23. Lowpass filter of the nth order. By taking the output signal from the currents in the resistors, a transfer function with both zeros and poles can be produced.

as a deterrent to the use of orders higher than about five for the filters described in this section.

Circuits based on the principle of fig. 21 can also be used to make elliptic filters. To obtain zeros in the transfer function, the input signal can again be taken to the output via two separate routes (see fig. 17). However, since every section now contains an amplifier, zeros can also be obtained in another way: by deriving the output signal from the *currents in the resistors of the filter*. This will be illustrated with a simple example, a filter of the second order as in fig. 19. The equations for the currents in the resistors are:

$$J_1 = V_i \frac{pC_1}{1 + p\tau_1 + p^2\tau_1\tau_2}, \qquad (19)$$

$$J_2 = V_i \frac{p^2C_2}{1 + p\tau_1 + p^2\tau_1\tau_2}. \qquad (20)$$

A circuit that gives a voltage proportional to the sum of these two currents will have a transfer function in which the polynomials of both numerator and denominator are functions of p; the transfer function therefore contains zeros.

A lowpass filter of this type and of order n is shown in *fig. 23*. At each place where there is a voltage follower in fig. 21, there are now *two* voltage followers, which use transistors in emitter-follower circuits. The emitter of one of these is connected to the resistor of the next

$$\frac{V_o}{V_i} = R_t \frac{R_0^{-1} + pC_1 + p^2\tau_1C_2 + \cdots}{1 + p\tau_1 + p^2\tau_1\tau_2 + \cdots}. \qquad (21)$$

Appropriate choice of the resistances and capacitances gives zeros in the stopband. Although the positions of the poles are determined only by the time constants (so that, in this respect the choice of the resistance and capacitance values remains free), the positioning of the zeros requires particular values of the capacitances and therefore of the resistances, as is clear from inspection of the numerator of (21). All available degrees of freedom of the passive elements have thus been used. If the required polynomial in the numerator has been established in the form

$$b_0 + b_1p + b_2p^2 + \ldots + b_np^n,$$

then the passive elements must be chosen so that

$$R_0 = 1/b_0; \; C_1 = b_1, \; C_2 = b_2/\tau_1, \ldots C_n = b_n/\tau_{n-1}.$$

The positioning of the zeros does not therefore require very much extra calculation.

Precision filters

We now return to a subject discussed earlier that can be important in choosing a filter, the parameter sensitivity. In some cases the specifications relating to parameter sensitivity can be very exacting. This is the case

in systems where a large number of filters are used in cascade, as in telephony. Because errors in the filters have a cumulative effect on the signals to be processed, particular attention must be paid to the accuracy of each filter.

We have stated earlier that low-loss elements are very important in the design of active filters. Since, in passive filters, the losses mainly originate in the inductors, precision filters can be made if circuits are available to simulate high-Q inductors. In this case the circuits of passive filters are used in which only the inductors are simulated and not complete circuits including inductors, as before. The simulation of inductors is made possible by the use of gyrators.

The gyrator

The ideal gyrator is a network element that was defined in 1948 by B. D. H. Tellegen [15]. It is a lossless two-port with the following relationships between the input and the output signals:

$$v_1 = - Ri_0, \qquad v_0 = Ri_1, \qquad (22)$$

The quantity R is the 'gyration resistance'. As in an ideal transformer, the sum of the powers at input and output of an ideal gyrator is always zero: these network elements are 'instantaneously lossless'. The symbol for a gyrator is shown in *fig. 24*.

From (22) it can be seen that the input voltage is proportional to the output current and the output voltage is proportional to the input current. It also follows from (22) that connecting an impedance Z_u across the output gives an input impedance equal to R^2/Z_u. If Z_u is due to a capacitance, i.e. $Z_u = 1/j\omega C$, then the input impedance becomes $j\omega R^2 C$. In that case the circuit behaves as an inductance

$$L = R^2 C. \qquad (23)$$

If $R = 1 \text{ M}\Omega$ and $C = 1 \text{ }\mu\text{F}$, then $L = 1 \text{ MH}$. (A conventional inductor of this inductance and with corresponding losses would weigh more than 1000 tons!)

Unlike all other passive circuit elements, the gyrator does not exhibit reciprocity. The gyrator cannot therefore be built up from combinations of other passive elements — resistors, capacitors, inductors and transformers. However, it is possible to produce gyrators from resistors and *active* elements [16]. The availability of integrated circuits has been an important stimulus to the construction and application of gyrators.

A circuit that works as a gyrator is shown in *fig. 25* (only the essential signal-processing parts are shown). It is formed by two differential amplifiers connected in parallel opposition. One of the stages contains two *N-P-N* transistors, the other two *P-N-P* transistors. For the circuit to approximate to the behaviour of an

ideal gyrator, the transistors must fulfil two conditions. Firstly, the current gain should be very large, e.g. > 1000; the current i_1 is then virtually equal to the current through R_2 and i_0 is virtually equal to the current in R_1. Secondly, the signal voltage between the base and the emitter should be small compared with the voltages across R_1 and R_2. The input voltage v_1 can then be equated to the voltage across R_1 and the output voltage v_0 equated to the voltage across R_2. To
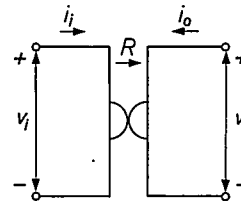
Fig. 24. Symbol for a gyrator. For such a circuit element $v_1 = - Ri_0$ and $v_0 = Ri_1$; the resistance R is the gyration resistance.
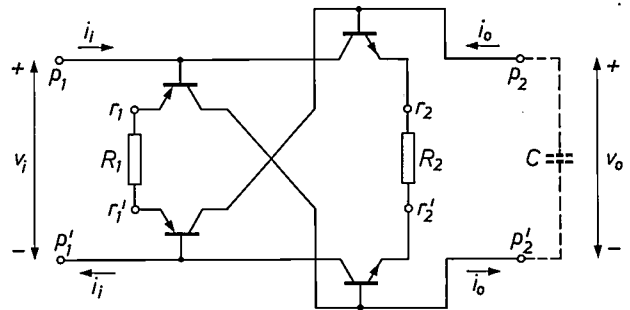
Fig. 25. Gyrator formed by two differential amplifiers connected in parallel opposition and two resistances R_1 and R_2. If a capacitor is connected across the terminals p_2 and p_2', the impedance across p_1 and p_1' corresponds to that of an inductance.

fulfil this condition, the transconductance of the transistors — the ratio of a.c. collector current to a.c. base-emitter voltage — must be large with respect to the reciprocal values of R_1 and R_2. If these two resistances are each 10 kΩ, then the transconductance of the transistors must be much larger than 0.1 mA/V, say 100 mA/V. With equal resistances ($R_1 = R_2 = R$) equations (22) are satisfied and we therefore have a gyrator.

Closer approximation to both conditions is obtained if each transistor is replaced by a number of transistors in a circuit with negative feedback. *Fig. 26* shows such a circuit, which simulates an *N-P-N* transistor. Base, collector and emitter connections are indicated by the letters b, c and e. If the current-gain factors β_n for the *N-P-N* transistors and β_p for the *P-N-P* transistor are

[15] B. D. H. Tellegen, Philips Res. Repts. 3, 81, 1948; see also B. D. H. Tellegen, Philips tech. Rev. 18, 120, 1956/57.

[16] There are also some very important gyrator devices that are not built up from conventional electronic components; these make use of gyromagnetic effects in ferromagnetic materials or piezoelectric effects.

large compared with unity, analysis of the circuit shows that the current gain β_n' of the simulation transistor is

$$\beta_n' \approx \beta_n{}^3\beta_p.$$

To make a simulated *P-N-P* transistor of very high current gain a different circuit is used. The reason for this is that with present integration techniques *N-P-N* transistors are fabricated in a 'vertical' configuration, which permits much higher current-gain factors than those of *P-N-P* transistors which are of 'lateral' configuration [17]. For example, β_n is usually between 50 and 150 whereas β_p lies between 3 and 10.

A circuit that simulates a high-gain *P-N-P* transistor is shown in *fig. 27*. The current-gain factor β_p' is given by

$$\beta_p' \approx \beta_n{}^2.$$

If $\beta_n = 100$ and $\beta_p = 5$ we find:

$$\beta_n' \approx 5 \times 10^6 \quad \text{and} \quad \beta_p' \approx 10^4.$$

For both simulation transistors, the transconductance S' is β_n times the transconductance S of the single transistor. The latter is given by qI/kT, where q is the electronic charge, I the collector current, k Boltzmann's constant and T the absolute temperature. For a current of 100 μA, the value of S is about 4 mA/V; taking, again, $\beta_n = 100$, the transconductance of the simulation transistors becomes:

$$S' = \beta_n S = 400 \text{ mA/V}.$$

If a gyrator with the above values of β_n', β_p' and S' is terminated by a capacitance C, then the simulated inductance is given by (23) to an accuracy better than 0.2%. The precision of the simulated inductor is therefore entirely determined by the precision of the gyration resistance and that of the capacitance. The rest of the circuit does not need to be of high precision; it is constructed as a monolithic integrated circuit on a crystal chip of area 3.4 mm². *Fig. 28* shows the circuit. The two resistors and the capacitors are connected externally [18].

In fig. 28 the dashed boxes delineate the four simulation transistors. The other transistors form d.c. sources of high internal resistance; this type of supply opens the possibility of making the input terminals 'floating' (neither input terminal connected to earth). The resistances R_1 and R_2 are connected between the points r_1, r_1' and r_2, r_2' (see also fig. 25). The input terminals are p_1, p_1' and the output terminals p_2, p_2'. The capacitor is connected across p_2 and p_2'. The diodes in the circuit serve to suppress parasitic effects. *Fig. 29* gives an impression of the size of a simulated inductor of inductance 1 MH. It consists of a gyrator and a capacitor of 1 μF.
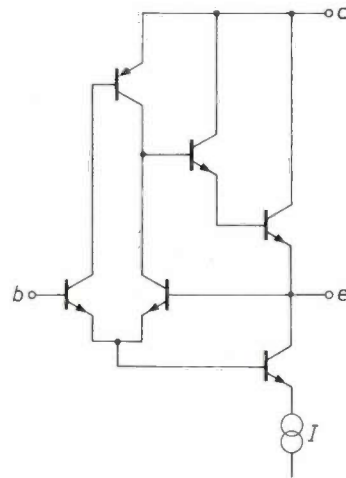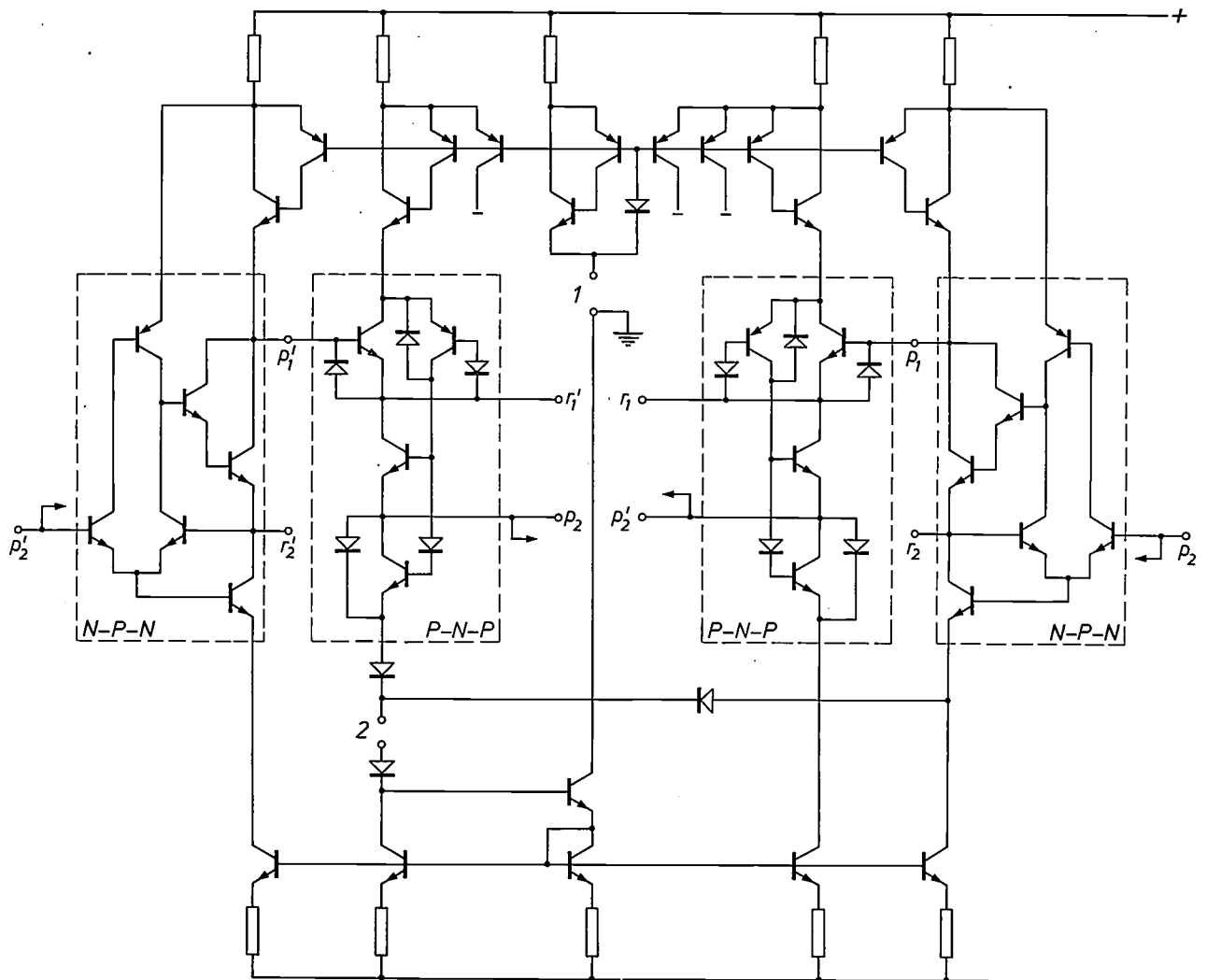

Fig. 26. Simulated *N-P-N* transistor. This circuit functions as an *N-P-N* transistor of high transconductance and current gain. Base, collector and emitter connections are indicated by *b*, *c* and *e*. *I* d.c. source.
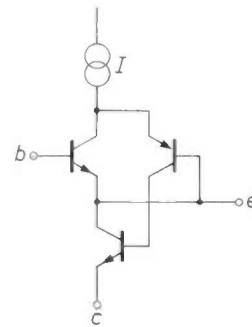
Fig. 27. Simulated *P-N-P* transistor. The letters *b*, *c*, *e* and *I* have the same significance as in fig. 26.
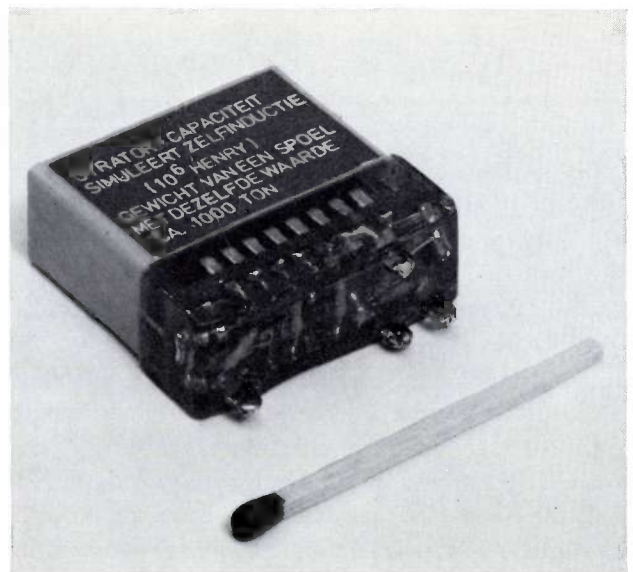
Fig. 29. Simulated inductor consisting of a gyrator as in fig. 28 and a 1 μF capacitor. With gyration resistances of 1 MΩ, the simulated inductance has the rather astonishing value of 1 MH.

Fig. 28. Circuit of a monolithic integrated gyrator. The two gyration resistances are external and are connected at r_1,r_1' and r_2,r_2' (see fig. 25). The simulated N-P-N and P-N-P transistors are enclosed by dashed boxes. The other transistors form d.c. sources of high internal resistance. The supply voltage is 12 V; if the gyration resistances are 10 kΩ, the whole circuit dissipates about 10 mW. Signals with peak amplitudes up to 1.6 V and 0.16 mA can be processed without distortion. The precision with which an inductor is simulated depends almost entirely on the precision of the gyration resistances and the capacitor connected at p_2,p_2'. The simulated inductance corresponds to the theoretical value given by (23) to within 0.2%. A resistance that determines the values of the direct current is connected between the terminals *1*. The terminals *2* can be either short-circuited or connected via one or more diodes; the number of diodes determines the maximum permissible drive voltage.

Filters with gyrators

With the gyrator shown in fig. 28, an inductor of very high Q can be simulated. If a capacitor C is connected across both input and output terminals, a parallel-resonant circuit of resonant frequency $1/RC$ is formed. At low frequencies the Q is greater than 500 and it increases with increasing frequency. Gyrators of this type can be used in classical LC filters which can be designed from existing tables. *Fig. 30* shows the circuit of a lowpass filter of this type for telephony [19]. It is made from ten capacitors and three gyrators. The passband extends to 3400 Hz. The attenuation is shown in *fig. 31* as a function of frequency. This characteristic can be obtained without the need of any adjustment of circuit parameters, using resistors and capaci-
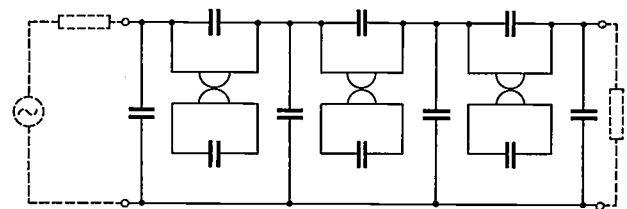
Fig. 30. Lowpass filter consisting of ten capacitors and three gyrators. If such a filter is supplied from an a.c. source of internal resistance equal to its resistive load, the transfer characteristic in the passband is little affected by parameter variations.

[17] See for example A. Schmitz, Solid circuits, Philips tech. Rev. 27, 192-199, 1966.

[18] For further details see J. O. Voorman and A. Biesheuvel, An electronic gyrator, IEEE J. SC-7, 469-474, 1972.

[19] R. Saal, Der Entwurf von Filtern mit Hilfe des Kataloges normierter Tiefpässe, Telefunken, Backnang 1961.

tors with a tolerance of 1 %. The reduction in volume achieved by the use of gyrators is demonstrated in *fig. 32*. This photograph depicts a filter of the type shown in fig. 30, together with a filter with the same

characteristics but containing inductors. The volume of the gyrator filter is about a quarter of that of the filter with inductors. At lower frequencies a much greater reduction in volume can be obtained.
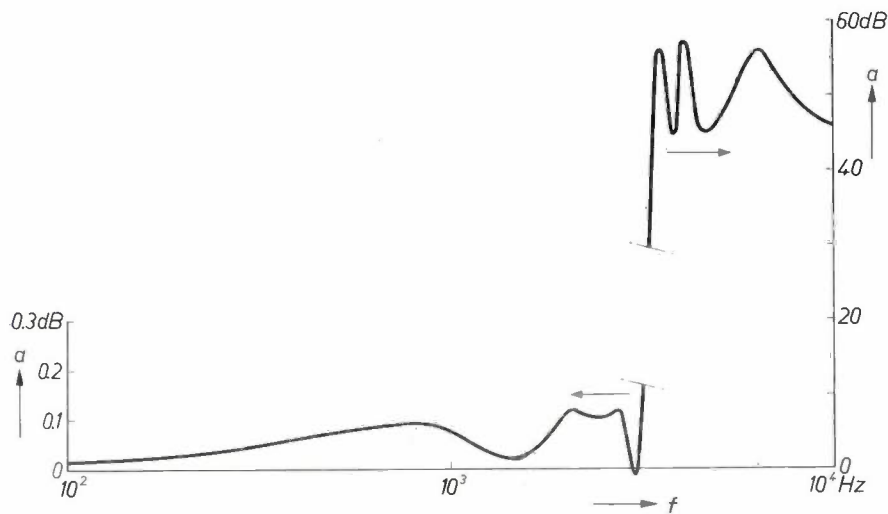
Fig. 31. The attenuation a of the filter shown in fig. 30 as a function of the frequency f.
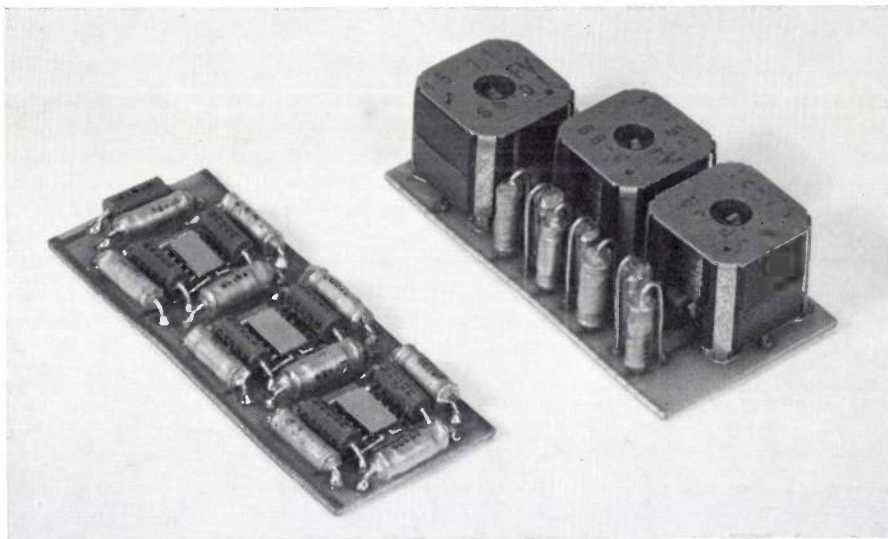
Fig. 32. Equivalent filters with gyrators (*left*) and with coils (*right*). The gyrator filter occupies only a quarter of the volume of the filter with inductors.

Summary. Conventional filters have always been made with resistors, capacitors and inductors. The current trend for miniaturization of electronic equipment has led to the design of filters with no inductors. This has been made possible by the introduction of one or more active elements. In this article three types of circuit for this purpose are described. In one of these, using only one amplifier, filters with one of the standard characteristics can be designed very simply with the aid of published tables and graphs. In the second type of filter, for which more than one amplifier is required, such tables are not required; the method used is well adapted to the synthesis of filters with an arbitrary transfer function. In active filters the amplitude of the signals must be limited to values that do not cause overload of the amplifiers. Voltage-magnification effects in the filter, which increase with the order of the filter, can make it difficult to use this method for the design of filters of higher order than five. Another point of interest in the choice of a filter is the effect of the parameter sensitivity on the shape of the transfer characteristic. This parameter sensitivity is very small in filters of the third type, in which low-loss inductors are simulated by means of gyrators produced by using transistors and diodes in integrated circuits. Gyrators permit the use of classical filter circuits and their existing design tables. Precision filters considerably smaller than classical filters with inductors can be made in this way.

Waves

Electromagnetic waves creeping through a metal at a few centimetres per second, elastic waves amplified by d.c. currents or which themselves generate d.c. currents, elastic surface waves launched and detected electrically — these are some of the remarkable phenomena that have come to light during the last 15 years in the world of electronics and solid-state physics. In the present issue of this journal they are discussed within the framework of a survey of various kinds of waves that are possible in insulating, conducting or piezoelectric solids.

A number of phenomena ordinarily not thought of as waves will be found to have a wave character. With the transverse electromagnetic waves, for example, we encounter not only light but also the skin effect; with the longitudinal electric waves we have not only plasma waves but also dielectric relaxation and Debye-Hückel screening.

Much attention is given to piezoelectric materials, in which elastic and electromagnetic waves are not independent of each other. In these materials electromagnetic waves (light) and elastic waves (sound) combine to give 'near-light', 'near-sound' and 'stiffened sound'. The stiffening is complex in piezoelectric semiconductors; when such a material carries a sufficiently large current, acoustic waves can be amplified.

The article concludes with a discussion of the remarkable Bleustein-Gulyaev wave — an electro-elastic surface wave which is a surface wave only by virtue of the piezoelectric nature of the substrate. The Bleustein-Gulyaev wave is presented as an 'imaginary' variant of a wave of 'stiffened sound' reflected at a surface.

Contents

Electromagnetic, elastic and electro-elastic waves

C. A. A. J. Greebe

Introduction

Electronics is very much concerned with electromagnetic waves — their generation, modulation, propagation, reception and processing. Electromagnetic waves may be transmitted through space or they may be guided by wires or other types of transmission line. Sometimes, especially in the microwave region, electromagnetic waves appear in the generating or processing equipment itself. Two important examples are the resonant cavity — where energy can be stored in the form of standing waves — and the delay line — where information can be stored in the form of modulated travelling waves.

Electronics also makes use of elastic waves: the quartz-crystal resonator is a very early and well known example. The use of elastic waves offers in many cases two notable advantages: the velocity of propagation is some $10^5 \times$ smaller than that of electromagnetic waves — so that, in 1 cm of a solid, elastic waves are delayed by the same amount as electromagnetic waves in 1 km of a cable; also, in certain carefully prepared materials, the attenuation of elastic waves can be relatively very small.

Elastic waves in solids are almost always generated and detected electrically. The conversion of electric signals into mechanical signals and vice versa is usually done by means of piezoelectric materials such as quartz; sometimes magnetostrictive materials are used.

Attempts to generate high-frequency elastic waves were for a long time limited to frequencies below 100 MHz because the electromechanical conversion was always done with *mechanically resonant* transducers. Such transducers must be only one or a few half wavelengths in thickness and above 100 MHz they became so thin as to be difficult to make or too fragile for practical use. This difficulty was surmounted during the fifties [1] and progress was such that some years later (1966) it was possible to generate and detect coherent waves of no less than 114 GHz [2]. One of the features of this breakthrough was the integration of transducer and medium: for example, elastic waves in a quartz crystal were generated and detected by virtue of the piezoelectric property of the crystal itself [2].

The use of elastic waves in electronics only really got under way after another development: the application of elastic *surface waves* [3]. As the name implies, these waves propagate only on the surface, leaving the bulk of the solid undisturbed. Like elastic waves in the bulk material they have a low velocity and, for well prepared surfaces, a low attenuation. They have however a great extra advantage: they are accessible over the whole length of their trajectory. This unique property opens up a whole range of possibilities which are easy to put into practice when the substrate is piezoelectric. The waves can then be generated, processed and detected by means of simple comb-shaped surface electrodes (interdigital transducers, see *fig. 1*); for example, filters with a wide range of characteristics can be made simply by choosing the shape, spacing and
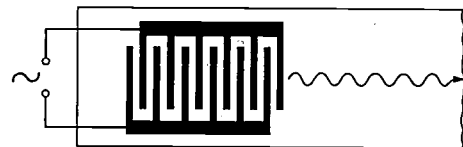

Fig. 1. Interdigital electrodes on a slice of a piezoelectric material (interdigital transducer) for the generation of elastic surface waves. The temporal frequency of the applied a.c. voltage, and the spatial frequency of the 'fingers' must correspond to the frequency and the wave number of the wave to be excited.

number of the 'teeth' of the electrodes [4]. Layer structures on the medium can be used to guide the waves or to give local changes in their dispersion. Delay lines based on surface waves can be provided with a large number of points where the signal may be tapped off [5]. The waves can be amplified by drift electrons in an adjacent semiconductor [6]. Finally, surface waves are particularly well adapted to systems of planar integrated circuits.

Dr C. A. A. J. Greebe is with Philips Research Laboratories, Eindhoven, and Professor Extraordinary at Eindhoven University of Technology. Prof. Greebe is now Director of the Institute for Perception Research (IPO), Eindhoven.

[1] H. E. Bömmel and K. Dransfeld, Phys. Rev. Letters **1**, 234, 1958 and **2**, 298, 1959, and Phys. Rev. **117**, 1245, 1960.
[2] J. Ilukor and E. H. Jacobsen, Science **153**, 1113, 1966.
[3] A survey is given in: R. M. White, Surface elastic waves, Proc. IEEE **58**, 1238-1276, 1970.
[4] See for example J. H. Collins and P. J. Hagon, Electronics **42**, No. 23, 97, 10 Nov. 1969, and R. F. Mitchell, Philips tech. Rev. **32**, 179, 1971.
[5] See for example J. H. Collins and P. J. Hagon, Electronics **43**, No. 2, 110, 19 Jan. 1970.
[6] J. H. Collins, K. M. Lakin, C. F. Quate and H. J. Shaw, Appl. Phys. Letters **13**, 314, 1968. See also J. H. Collins and P. J. Hagon, Electronics **42**, No. 25, 102, 8 Dec. 1969.

In recent years there has therefore been a growing interest in all sorts of wave phenomena — bulk waves and surface waves, electromagnetic waves and elastic waves and combinations of these in piezoelectric materials. It seemed to be useful to attempt a systematic review of these various forms against a background of conventional, well known forms of wave propagation. This article, therefore, is meant as a sort of 'introduction to waves, and is of a tutorial nature; it gives no scientific 'news' but presents known material and points out relationships. The opportunity will also be taken of discussing certain perhaps in practice less important but nevertheless remarkable wave phenomena such as helicon waves.

In the first part of the article we shall consider wave propagation in *unbounded* media — for example in free space, in optically anisotropic media and in conductors with and without a magnetic field — starting from the differential equations for the appropriate variables of the medium. The travelling waves that we find are characterized by an angular frequency ω ($2\pi \times$ the frequency) and a wave vector k (whose components k_x, k_y and k_z are respectively 2π divided by the wavelengths in the x-, y- and z-directions). The waves may grow or diminish in both space and time (see fig. 3, p. . . .), which is indicated by k or ω having an imaginary part. A very important aspect of a wave phenomenon is the *dispersion relation* which is the relation between ω and k. Among the subjects dealt with in this first part of the article are the familiar waves of light and sound; the strongly attenuated propagation in metals resulting in the skin effect; a variant of this in a strong magnetic field, the 'helicon' waves, and some longitudinal electric waves. We shall also consider a situation where the wave does not propagate in the direction of the wave vector k, a matter to be borne in mind when considering anisotropic materials, such as crystals, whether carrying bulk or surface waves.

The second part of the article deals with the coupling of waves in unbounded media. Wave propagation in piezoelectric materials can be very complicated because the electric and elastic variables are not independent of each other. If, however, the coupling is weak, the problem can be considerably simplified by regarding the waves as coupled electric and elastic waves, each of which would propagate independently if the coupling were in fact zero. This method of attack can also be useful in other cases where there are many coupled variables. Among the examples discussed here is the amplification of acoustic waves ('acoustic amplifier').

In the third part of the article combinations of waves that can exist in two adjacent media are discussed.

These include combinations of incident, refracted and reflected waves and also — our particular concern here — surface waves. A surface wave occurs in the well known phenomenon of total internal reflection, but in this case it occurs only in combination with the incident wave and the reflected wave. Modern developments in electronics, however, are concerned with true surface waves that are independent of any bulk waves. A simple example — the Bleustein-Gulyaev wave — will be discussed at length.

In concluding this introduction attention should be directed to a problem that will not be dealt with in this article but is of the greatest importance to investigations of wave behaviour in unusual, novel media. In general a travelling wave transports energy of which, usually, a fraction is dissipated in the medium. For a given real frequency the wave amplitude then diminishes in the direction of propagation (k is partly imaginary); the medium is passive. There are, however, media which can be activated in one way or another; in such media waves are possible that become larger in the direction of propagation. In the acoustic amplifier, for example, the medium is a piezoelectric semiconductor which is fed with energy by means of a d.c. current; this energy is partly taken up by the acoustic wave. In a well designed device, the input signal re-appears, after traversing the medium, amplified at the output. It is, however, not at all certain that a medium in which such 'amplifying' waves are theoretically possible will necessarily be potentially useful as an amplifier. It is possible, for example, that the medium will exhibit 'absolute instabilities' and reacts to an input signal with an explosive increase of the variables. In this case the output signal is no longer related in any way to the input signal. A. Bers and R. J. Briggs have given a theoretical approach to the problem of how to decide, on the basis of the dispersion relation, whether a new medium will have absolute instabilities or whether it can be used for amplification [7]. The investigation of how the medium reacts to an excitation (input signal, source) is inherent to this analysis. We shall leave this question completely aside and consider only freely propagating waves, without enquiring how they are generated.

[7]   See R. J. Briggs, Electron-stream interaction with plasmas, M.I.T. Press, Cambridge (Mass.) 1964, chapter 2.

Waves in unbounded homogeneous media

Main features of the analysis

The method of analysis of wave phenomena in unbounded media will be illustrated by means of a simple one-dimensional example: an infinitely long uniform transmission line with a capacitance of C(F/m) per unit length and an inductance of L(H/m) per unit length; see *fig. 2*. Changes in current in this transmission line give rise to voltage differences across the inductances. The capacitances are charged by the difference in the currents before and after them, so that the voltages across the capacitances also change.
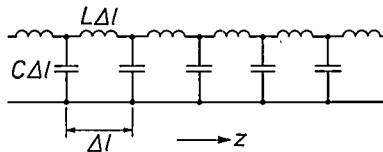
Fig. 2. Transmission line with a shunt capacitance $C\Delta l$ and a series inductance $L\Delta l$ per section Δl. When $\Delta l \to 0$, with C and L remaining constant, we get a uniform continuous line as discussed in the text.

These qualitative relations between the two wave variables of this problem, the voltage V and the current I can be quantified in two homogeneous linear differential equations in the spatial coordinate z and the time t:

$$\frac{\partial V}{\partial z} + L\frac{\partial I}{\partial t} = 0,$$
$$C\frac{\partial V}{\partial t} + \frac{\partial I}{\partial z} = 0. \tag{1}$$

The solutions of these equations are exponential functions of z and t:

$$V = V_0 \exp \mathrm{j}(\omega t - kz),$$
$$I = I_0 \exp \mathrm{j}(\omega t - kz). \tag{2}$$

All the waves discussed in this article can be described as functions of this form. The real parts of such complex expressions represent the actual physical quantities. If ω and k are real, as we shall provisionally assume, we have waves in their simplest form: sinusoidal functions of position which propagate at the velocity ω/k, the *phase velocity*. The amplitudes V_0 and I_0 can be complex; the modulus of the complex amplitude is what we normally call the amplitude whilst its argument gives the phase of the disturbance.

Substituting (2) in (1) yields two homogeneous linear equations for the complex amplitudes:

$$kV_0 - \omega LI_0 = 0,$$
$$\omega CV_0 - kI_0 = 0. \tag{3}$$

There are solutions to (1) only when the determinant of the coefficients of (3) is zero and this condition gives the dispersion relation

$$k^2 - \omega^2 LC = 0. \tag{4}$$

From this we derive the phase velocity

$$v = \omega/k = \pm 1/\sqrt{LC},$$

which, combined with (3), gives the following ratio of the complex amplitudes:

$$V_0/I_0 = \pm \sqrt{L/C}. \tag{5}$$

The positive root is called the *characteristic impedance* of the transmission line.

The steps outlined above are typical of many problems of wave motion. We shall always express the properties of the *medium* or the *physical system* in terms of differential equations in the *wave variables*. We shall restrict ourselves, as above, to homogeneous linear differential equations, whose coefficients are independent of time and place: this expresses the fact that the properties of the medium remain constant and are spatially homogeneous. Substitution of *harmonic* waves leads to homogeneous linear algebraic equations for the complex amplitudes. In a well formulated problem, the number of these equations is equal to the number of variables. Putting the determinant of these equations equal to zero yields the *dispersion relation*. Subsequently, we can in general calculate all the complex amplitudes in terms of one of them and so find the ratios of all the real amplitudes as well as all the phase differences — that is to say, the 'structure' of the wave.

If there are several harmonic solutions these can be quite freely superposed. Superposition implies, by its nature, that the behaviour of each component wave is entirely independent of the presence of the others: there is no interaction between the components. The situation is quite different when the differential equations contain nonlinear terms. If such terms are sufficiently small, it is often possible to consider a solution as the sum of several approximately harmonic components, but the behaviour of each component will now depend on the presence of the other components: the components interact, some becoming stronger, others weaker.

The dispersion equation; dispersion

The left-hand side of the dispersion relation (4), i.e. the determinant of (3), can be factorized into two factors. If one of these is set equal to zero we get the dispersion relation for one type of wave, e.g. a wave travelling to the left ($v < 0$). The other factor set equal to zero gives a wave travelling to the right ($v > 0$).

This is a trivial example of what one always tries to do: to resolve the determinant of the wave problem into factors — setting each factor equal to zero gives a dispersion relation for one type of wave. A less trivial example is found in the problem of sound waves in an isotropic solid: when the equations are set up sufficiently generally, two factors are found in the determinant, one corresponding to longitudinal waves and the other to transverse waves. It is also possible to reverse this whole approach. For example, in this article we shall assume — to stay with sound waves in solids — a longitudinal wave in the z-direction and find out for which combinations of ω and k this is possible. What we then find is the dispersion relation for longitudinal waves in the z-direction in the given medium, and the structure of these waves. In the case of an isotropic substance, the characteristics of longitudinal waves in any direction would then also be known. However, whether other waves could exist in the medium then remains an open question.

In the transmission line all harmonic waves travelling to the right have the same velocity v. If a disturbance consists only of waves travelling in this direction, therefore, these all continue to proceed *together* along the line, i.e. they do not *disperse* from one another, so that the disturbance or signal *retains its form* while propagating at a velocity v to the right. The transmission line is then called a *dispersionless* system. We shall encounter many other dispersionless systems but also systems *with* dispersion in which v is a function of k and where the shape of a disturbance in general changes as it is propagated.

Complex wave number and complex frequency

If the transmission line of fig. 2 has not only series inductance but also series resistance (R per unit length, in Ω/m), a term IR must be added to the left-hand side of the first equation (1). Repeating the procedure outlined above, we arrive at the dispersion relation

$$k^2 + j\omega RC - \omega^2 LC = 0.$$

Expressions (2) are thus no longer solutions for real ω and k. This is obvious physically: the line is no longer lossless so that the waves are attenuated as they are propagated. Our whole scheme can however still be

retained and the attenuation included if ω and k are allowed to be complex.

When ω and k are written as the sums of real and imaginary parts:

$$\omega = \omega_r + j\omega_i,$$
$$k = k_r + jk_i,$$

the waveform (2) can be expressed as the product of an exponential and a harmonic factor:

$$\exp j(\omega t - kz) = \exp(-\omega_i t + k_i z)\exp j(\omega_r t - k_r z). \quad (6)$$

This represents a sinusoidal wave (the second factor) whose amplitude diminishes (or grows) both with time and from place to place. The general case is illustrated in the central curve of *fig. 3*. The other curves show the nature and behaviour of the excitation if ω or k is purely real or purely imaginary. All these and the intermediate cases can be regarded as kinds of wave. Among them are phenomena which in ordinary experience would not be called waves, for example the alternating field in a waveguide when this is excited at a (real) frequency below the cut-off frequency; k is then purely imaginary. Such a cut-off wave (or evanescent mode) is shown in fig. 3c.

Waves in three dimensions

For wave phenomena in three dimensions, the term kz in equation (2) must be replaced by $\mathbf{k} \cdot \mathbf{r}$, where $\mathbf{r}$ is the radius vector of a point in space defined by coordinates x, y, z, and $\mathbf{k}$ is the wave vector having the components k_x, k_y, k_z along these coordinates: $\mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z$. If $\mathbf{k}$ is complex it can be represented by the two real vectors $\mathbf{k_r}$ and $\mathbf{k_i}$:

$$\mathbf{k}(k_x, k_y, k_z) = \mathbf{k_r}(k_{rx}, k_{ry}, k_{rz}) + j\mathbf{k_i}(k_{ix}, k_{iy}, k_{iz}),$$
$$k_x = k_{rx} + jk_{ix},$$
$$k_y = k_{ry} + jk_{iy},$$
$$k_z = k_{rz} + jk_{iz}.$$

If $\mathbf{k_r}$ and $\mathbf{k_i}$ are parallel to one another, in other words, if the ratios k_{rx}/k_{ix}, k_{ry}/k_{iy}, k_{rz}/k_{iz} are equal, then the problem can be reduced once more to a one-dimensional one. We only have to rotate the coordinate system until the new z-axis coincides with the common direction of $\mathbf{k_r}$ and $\mathbf{k_i}$; then $\mathbf{k} \cdot \mathbf{r} = k_z z$. The waves are in this case essentially one-dimensional and *plane* waves: the wave variables are independent of the (new) x- and y-coordinates. The (new) x,y-planes (perpendicular to $\mathbf{k_r}$ and $\mathbf{k_i}$) are *wavefronts*.

When $\mathbf{k_r}$ and $\mathbf{k_i}$ are not parallel, the wave is essentially *not* one-dimensional. This is the case, for example, for surface waves, which are propagated parallel to the surface ($\mathbf{k_r}$ // surface) but usually fall off exponentially in the perpendicular direction ($\mathbf{k_i}$ ⊥ surface).

With more than one dimension there is still only one dispersion relation. This implies that a great variety of waves is possible since, of the four complex quantities k_x, k_y, k_z and ω, three are in general independent.

Notation

In order to avoid more indices than are really necessary, we shall usually make no distinction between a

The differential operators $\partial/\partial t$, $\partial/\partial x$, ... will be abbreviated to ∂_t, ∂_x, ... The algebraic equations of the type (3) are obtained from the differential equations of the type (1) by replacing the operator ∂_t by the factor $j\omega$, ∂_x by the factor $-jk_x$, etc.

We shall also be concerned below with curls and divergences of the vectorial wave variables. In terms of Cartesian coordinates the curl and divergence of an arbitrary vector a are defined as follows:
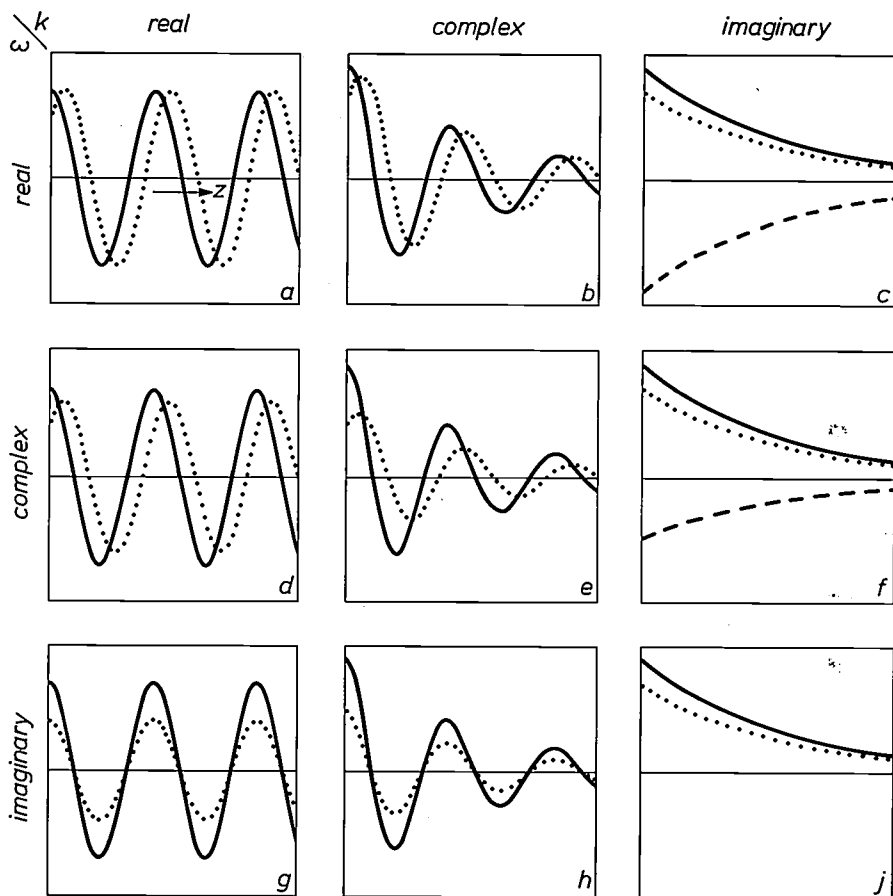

Fig. 3. The character of the various waves represented by the expression $\exp j(\omega t - kz)$, classified according to whether ω and k are real, imaginary or complex (see eq. 6). Solid curves: the waveform at a given instant. Dotted lines: a fraction of a period later. Dashed lines (c and f only): half a period later. Only cases (a), (b), (d) and (e) represent travelling waves in the conventional sense. It is assumed that ω_r and k_r have the same sign: the waves travel from left to right ($+z$-direction). It is assumed that ω_i is positive and k_i is negative, where they arise: the amplitudes decrease with time and from left to right. The opposite sign for ω_i or k_i would imply waves of increasing amplitude.

complex variable, its real part (i.e. the actual physical quantity) and the complex amplitude. In equations of the type (3) and (5) we shall therefore omit the indices 0. This should give no difficulties: where the distinction is important it is usually clear from the context what is meant. Some care may be necessary with nonlinear combinations and relations; an expression such as IV for power, for example, is correct only if I and V are the actual instantaneous current and voltage [8].

$$(\text{curl } a)_x = \partial_y a_z - \partial_z a_y,$$
$$(\text{curl } a)_y = \partial_z a_x - \partial_x a_z,$$
$$(\text{curl } a)_z = \partial_x a_y - \partial_y a_x,$$
$$\text{div } a = \partial_x a_x + \partial_y a_y + \partial_z a_z.$$

[8] The power is thus $(\text{Re} I)(\text{Re} V)$ in terms of complex variables I and V. Usually only the time average of such a product, $\overline{(\text{Re} I)(\text{Re} V)}$, is of interest; this is given by $\frac{1}{2}\text{Re}(IV^*)$, where the asterisk denotes a complex conjugate.

Electromagnetic waves

Maxwell's equations

Many natural phenomena are wholly or partly electromagnetic. The set of differential equations describing a wave phenomenon often, therefore, involves Maxwell's equations in one way or another. In their most general form these are, in SI units:

$$\text{curl } H = \dot{D} + J, \quad (7) \qquad \text{div } B = 0, \quad (9)$$

$$\text{curl } E = -\dot{B}, \quad (8) \qquad \text{div } D = \varrho_e, \quad (10)$$

where H is the magnetic field, E the electric field, B the magnetic flux density, D the dielectric displacement, J the current density and ϱ_e the charge density.

'Divergence' can be interpreted as 'strength of source'. Thus (10) states that charge is the source of the D field and (9) states that the B field has no sources. Similarly we can say that 'curl' is equivalent to 'vortex strength' [9].

If we take the divergence of (7), remembering that the div curl of any vector is zero, and combine the result with (10), we find the continuity equation for the charge:

$$\text{div } J = -\dot{\varrho}_e. \quad (11)$$

This equation states that the charge decreases at locations where there is a source of current.

Finally, there is an important energy equation:

$$-\text{div } [E \times H] = E \cdot J + E \cdot \dot{D} + H \cdot \dot{B}, \quad (12)$$

which is found by combining (7) and (8) with the vector identity $-\text{div } [a \times b] = a \text{ curl } b - b \text{ curl } a$. Equation (12) may be interpreted as follows: energy is transported by the electromagnetic field with an energy-flow density given by the Poynting vector [10]

$$S = E \times H. \quad (13)$$

The three terms on the right-hand side of (12) thus represent sinks (negative sources) for the energy flow. The first term represents the development of ohmic heat, the second the storage of electrical energy and dielectric losses, and the third the storage of magnetic energy and magnetic losses. We shall encounter the second term $E \cdot \dot{D}$ again in our considerations below.

Maxwell's equations are 'laws of nature' in the sense that they are always and everywhere valid. However, they leave a considerable freedom in the behaviour of the electromagnetic variables: they give only 8 scalar relations as against 16 scalar variables. The properties of the wave are further determined by the properties of the medium. Therefore one can expect new and unusual electromagnetic phenomena if new and unusual media become available. An example is furnished by the remarkable helicon waves, first discovered on paper, which can be generated in very pure sodium at very low temperatures (4 K) in a strong magnetic field (10^4 Oe). These electromagnetic waves, which will be discussed in more detail below, propagate with almost no attenuation at the unusually low velocity (for electromagnetic waves) of, say, 10 cm per second.

In simple cases the properties of the medium can be specified by three constants of the material: the permittivity ε, the magnetic permeability μ and the conductivity σ. The following three equations then describe the medium:

$$D = \varepsilon E, \quad B = \mu H, \quad J = \sigma E. \quad (14)$$

In the special case where the medium is free space,

$$D = \varepsilon_0 E, \quad B = \mu_0 H, \quad J = 0, \quad \varrho_e = 0. \quad (15)$$

When D and $\varepsilon_0 E$ differ, as they do for a physical medium, this is a consequence of the electric polarization of the medium. Equally, any difference between B and $\mu_0 H$ is a consequence of the magnetization of the medium.

If we introduce the electric polarization P via the definition $D = \varepsilon_0 E + P$, the expression $E\mathrm{d}D$ for the electrical energy delivered by the electromagnetic field in time dt (see the text referring to eq. (12) and (13)) becomes clearer physically: $E\mathrm{d}D = E\mathrm{d}(\varepsilon_0 E + P) = \mathrm{d}(\frac{1}{2}\varepsilon_0 E^2) + E\mathrm{d}P$. The first term is the increase in the free-space field energy and the second term is the work done on the medium by the field (force × displacement).

Taking the equations (14) together with (7), (8), (9) and (10), the medium seems to be 'overdetermined': we have five vectors D, J, E, B and H and one scalar ϱ_e and also five vector equations in (7), (8) and (14) but two scalar equations (9) and (10). However, the derivative of (9) with respect to time, div $\dot{B} = 0$, is a direct consequence of (8) (because div curl = 0). For our time-dependent waves, with $\dot{B} = \mathrm{j}\omega B$ and $\omega \neq 0$, this means that (8) implies (9). In a more general situation, the independent information given by (9) is concerned only with the constant (time-independent) part of B.

In what follows we shall first derive the velocity and the structure of electromagnetic waves in free space. Then we shall consider other non-conducting media ($\sigma = 0$). If in such media ε and μ are truly constants of the material, i.e. wholly determined by the medium and not at all by the wave, then we should find waves that are qualitatively the same as in free space. An interesting phenomenon that does not occur in free space, double refraction, can be related to anisotropy of the medium; to describe the medium in such a case, instead of the constant ε, six constants are necessary (in the worst case) combined in the permittivity tensor ε. Other phenomena that do not occur in free space, dispersion and absorption, can be described by a formal extension of the concept of permittivity to a

frequency-dependent complex permittivity (which thus also depends on the wave). Of course, this does not *explain* dispersion and absorption; to do this the required $\varepsilon(\omega)$ has to be related to the structure of the medium. Faraday rotation, as we shall see, can also be described formally in the same way, using a particular complex permittivity tensor.

The same methods can be used to describe wave propagation in conducting media, because the conductivity can be represented by an imaginary part of the permittivity. In this case we shall proceed less formally and derive for example an effective ε-tensor that describes the propagation of helicon waves on the basis of the behaviour of the conduction electrons in a strong magnetic field. We shall also encounter longitudinal electric waves which are not possible in free space because ε_0 is not zero, but which may occur in conductors under certain conditions when the *effective* permittivity is zero.

Electromagnetic waves in non-conducting media

Free space

For the analysis of electromagnetic waves in free space we start with Maxwell's equations, combined with the equations (15). (We omit here the suffix 0 from ε and μ; some of the results can then be used later.)

When we substitute $j\omega$ for ∂_t and assume non-zero ω, the equations become considerably simpler. In view of the identity div curl $\equiv 0$, not only does (9) follow from (8) but also (10) follows from (7) because J and ϱ_e are both zero. Two vector equations are thus left over for the two vectors E and H:

$$\begin{aligned} \text{curl } H &= \quad j\omega\varepsilon E, \\ \text{curl } E &= -j\omega\mu H. \end{aligned} \quad (16)$$

For plane waves propagating in the z-direction we have therefore (with $\partial_z = -jk$, $\partial_x = \partial_y = 0$):

$$\begin{aligned} \omega\varepsilon E_x - \quad kH_y &= 0 \\ kE_x - \omega\mu H_y &= 0 \end{aligned} \right\} (a)$$

$$\begin{aligned} \omega\varepsilon E_y + \quad kH_x &= 0 \\ kE_y + \omega\mu H_x &= 0 \end{aligned} \right\} (b)$$

$$\omega\varepsilon E_z = 0 \quad (c)$$

$$\omega\mu H_z = 0 \quad (d)$$

$$(17)$$

The determinant of these equations can be seen to factorize into four factors, and by putting each factor separately equal to zero we find in principle (see p. 314) four dispersion relations, each representing a wave. In each of the four waves, given by (a), (b), (c) and (d) the wave variables are different.

Now the pair of equations (a), with $\varepsilon = \varepsilon_0$ and $\mu = \mu_0$, already describe all the properties of electromagnetic waves in free space (see *fig. 4a*). Their velocity — the velocity of light in free space — follows from the dispersion relation for (a):

$$k^2 - \varepsilon_0\mu_0\omega^2 = 0, \quad (18)$$

and is therefore $1/\sqrt{\varepsilon_0\mu_0}$. The waves are transverse (E and H both perpendicular to k) and E and H are also perpendicular to each other. The ratio of the complex amplitudes E_x/H_y is equal to $k/\varepsilon_0\omega = \sqrt{\mu_0/\varepsilon_0}$, the *intrinsic impedance of free space*; and since this ratio is real, E_x and H_y are in phase. (The concept of the intrinsic impedance of a medium is directly analogous to that of the characteristic impedance of a line.)
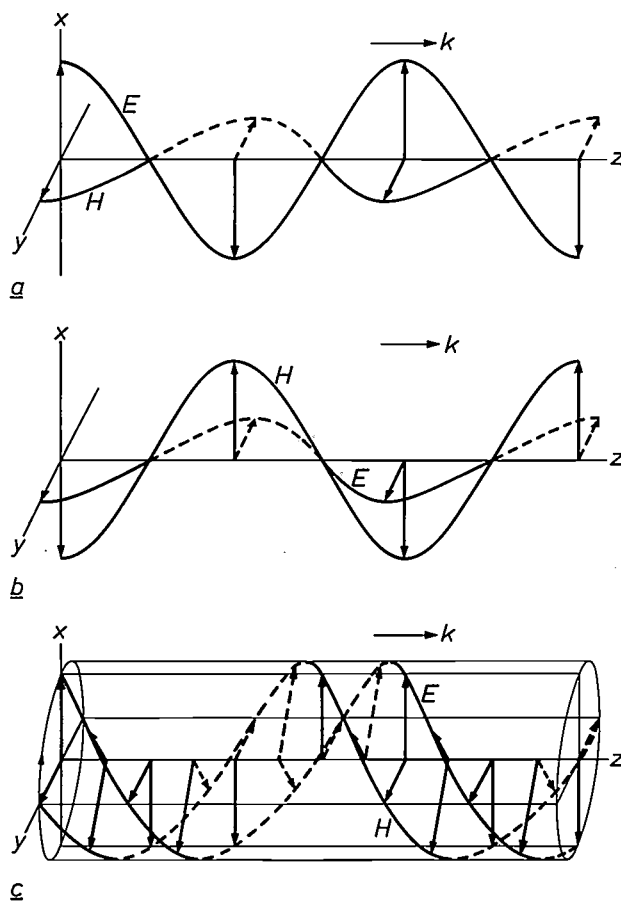

Fig. 4. Structure of electromagnetic waves in free space, *a*) corresponding to (17a), *b*) corresponding to (17b). The wave (*b*) is simply the wave (*a*) rotated through 90° about the z-axis. These waves are plane polarized. A circularly polarized wave (*c*) is obtained by superposition of waves (*a*) and (*b*) of equal amplitude but with 90° phase difference.

---

[9] See for example J. Volger, Vortices, Philips tech. Rev. 32, 247-258, 1971.

[10] In (13) E and H are the actual electric and magnetic fields. Using complex wave variables, the time-averaged Poynting vector is $\overline{S} = \frac{1}{2}\text{Re}[E \times H^*]$; cf. note [8].

The practical significance of the expression $1/\sqrt{\varepsilon_0\mu_0}$ for the velocity of light in free space is that, in the construction of a system of units such as SI, although there is some freedom of choice with respect to ε_0 and μ_0 the combination $1/\sqrt{\varepsilon_0\mu_0}$ must always be equal to the velocity of light.

Other solutions are obtained by applying a rotation to that of fig. 4a; k (or ω) and E (or H) can be freely chosen. The solution of the equations (17b) is simply the wave of (17a) rotated through 90° about the z-axis (fig. 4b). Any wave in free space can be described by a superposition of such waves.

One rather special case of superposition is the superposition of (a) and (b) where E_x in (a) and E_y in (b) are of the same amplitude but differ by 90° in phase:

$$E_x = \pm jE_y. \tag{19}$$

This is a circularly polarized wave (fig. 4c). The upper sign (+) represents a vector rotating clockwise, and the lower sign (−) represents a vector rotating anticlockwise, as seen by an observer looking in the +z-direction, and assuming ω to be positive [11].

The end points of the vectors lie on a helix. The relation between the sense of this helix, the sense of rotation of the vectors and the direction of wave propagation can best be formulated by adopting the convention used in optics. In this convention a sense of rotation is defined as that seen by an observer receiving the waves. Then the sense of the helix is the same as that of the rotation of the vectors (in whatever direction the wave propagates), and this is by definition the sense of the circular polarization. According to this definition the upper sign in (19) represents left-handed circular polarization for a wave travelling in the +z-direction and right-handed circular polarization for a wave travelling in the −z-direction.

The equations (17c) and (17d) would represent longitudinal electric waves and longitudinal magnetic waves
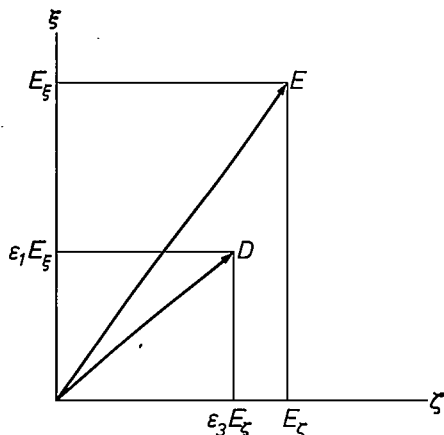

Fig. 5. If a material has a different polarizability in two directions ξ and ζ (see eq. 20), the ratios D_ξ/E_ξ and D_ζ/E_ζ are not equal. Hence D and E are no longer parallel to one another, unless they happen to be along the ξ-axis or the ζ-axis.

respectively, but their dispersion relations, $\omega\varepsilon = 0$, $\omega\mu = 0$, are not satisfied in free space (for $\omega \neq 0$). For this reason longitudinal electromagnetic waves cannot exist in free space.

For a medium whose electric and magnetic properties are described by (14), where ε and μ are true constants of the material and where σ is zero, electromagnetic waves entirely analogous to those in free-space waves are possible; their velocity is $1/\sqrt{\varepsilon\mu}$ and the intrinsic impedance E/H is $\sqrt{\mu/\varepsilon}$. Such media do not really exist but in certain cases — for example that of low-frequency waves in an isotropic lossless insulator — the wave propagation is well described in this way.

We shall now examine what happens when the medium is not isotropic.

Anisotropy; double refraction

Let us consider a crystal that is not equally polarizable in all directions. The relation between D and E can now no longer be characterized by a single scalar quantity. We shall assume that an orthogonal coordinate system ξ, η, ζ exists in which

$$D_\xi = \varepsilon_1E_\xi, \quad D_\eta = \varepsilon_1E_\eta, \quad D_\zeta = \varepsilon_3E_\zeta, \tag{20}$$

where $\varepsilon_3 > \varepsilon_1$; the polarizability is thus larger in the ζ-direction than in the ξ,η-plane. We are then concerned with a uniaxial crystal in which the ζ-axis is the optical axis. From (20) we can immediately conclude — see *fig. 5* — that D and E are *no longer parallel* to each other, unless they happen to be parallel or perpendicular to the ζ-axis.

For light propagated along the optical axis, the calculation of p. 317 can again be used, with $\varepsilon = \varepsilon_1$. For propagation perpendicular to the optical axis, the calculation is also entirely analogous, except that in (17a), $\varepsilon = \varepsilon_3$ and in (17b), $\varepsilon = \varepsilon_1$: we therefore get *two waves*, one polarized along the optical axis and the other perpendicular to it, with *different velocities*.

The situation becomes really interesting when we consider plane waves whose k vector makes an angle other than 90° with the ζ-axis. Going back to Maxwell's equations (7) and (8) we find, taking a coordinate system x,y,z in which k is parallel to the z-axis, and taking $B = \mu H$, $J = 0$, $\mu \neq 0$ and $\omega \neq 0$:

$$
\begin{aligned}
kH_y &= \omega D_x, & \omega\mu H_x &= -kE_y, \\
kH_x &= -\omega D_y, & \omega\mu H_y &= kE_x, \\
0 &= D_z, & H_z &= 0.
\end{aligned}
\tag{21}
$$

It follows that D, H and k are perpendicular to one another. D and H are still transverse: the wavefronts are D,H-planes. If D is taken perpendicular to the ζ-axis (fig. 6a) the situation is still quite unremarkable: E is again parallel to D and the wave has the same nature as we have already encountered. If, however, D

is taken to be in the k,ζ-plane (fig. 6b) then D is neither parallel nor perpendicular to the ζ-axis, so that E is *no longer parallel* to D (see fig. 5). The *Poynting vector* $S = E \times H$ is therefore *no longer parallel* to the *wave vector k*. Since S gives the direction of the energy flow and therefore, in the case of a parallel beam, the direction of the beam, the wavefronts lie obliquely to this direction (*fig. 7*). An unpolarized beam falling perpendicularly on the x,y-plane in fig. 6a and b (this being the surface of the crystal) is therefore split into a 'straight-through' beam (fig. 6a) and an 'oblique' beam (fig. 6b): this is double refraction.

The permittivity tensor

We have seen that in an anisotropic material D and E are in general not parallel to one another (see fig. 5). It follows that, in a coordinate system x,y,z not parallel to the coordinate system ξ,η,ζ, used above, each

Fig. 7. A beam, i.e. a wave with bounded wavefronts, is propagated in the direction of the Poynting vector S. If k is not parallel to S (fig. 6b), the wavefronts are not perpendicular to the beam.
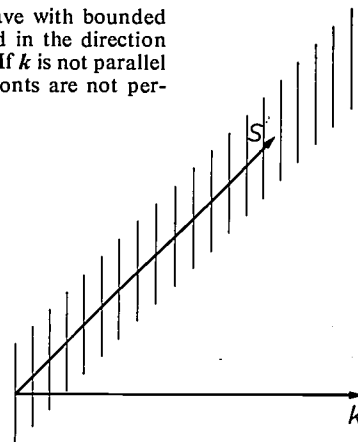
component of D may depend on all the components of E. In general we must write:

$$D_k = \sum \varepsilon_{kl} E_l. \quad (k,l = x,y,z) \tag{22}$$

This applies also to a biaxial crystal for which *three* different ε's occur in (20). The two or three ε's in (20) and the nine ε_{kl}'s in (22) give *the same linear relationship* between the physical vector fields D and E; if yet another coordinate system is chosen, the nine quantities describing this relationship will have other values. Such a *tensor relation* between D and E is written:

$$D = \varepsilon E, \tag{23}$$

where ε, the *permittivity tensor*, is thus a property of the crystal which, for each coordinate system x,y,z, is defined by another array

$$\begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix} \tag{24}$$

of nine scalar quantities.

In such a crystal the electric state and therefore the electric energy per unit volume U_E are determined by the values of E_x, E_y and E_z. In a change of state in which D increases by dD, the crystal takes up an energy of $E \cdot dD$ per unit volume from the field (see p. 316) and U_E thus changes by this amount so that

$$dU_E = E \cdot dD = \sum_{k,l} \varepsilon_{kl} E_k dE_l.$$

It follows that

$$\partial U_E / \partial E_l = \sum_k \varepsilon_{kl} E_k,$$

and

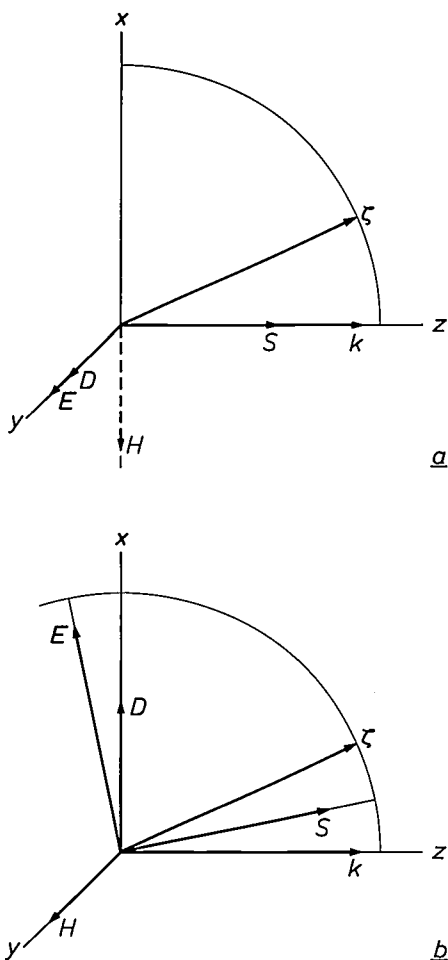$$\partial^2 U_E / \partial E_k \partial E_l = \varepsilon_{kl}.$$

Fig. 6. Double refraction. In a uniaxial crystal, a plane wave whose k vector makes an angle with the optical axis ζ, the vectors D, H and k (see eq. 21) are still perpendicular to one another, as in an isotropic medium. If D is perpendicular to the optical axis as in (a), then E is parallel to D and the Poynting vector S is parallel to k; if, however, D lies in the plane defined by k and ζ as in (b), then E is not parallel to D and so S is not parallel to k.

[11] A right-handed coordinate system is assumed here as in fig. 4. It is further assumed that all the wave variables are proportional to exp $(+ j\omega t)$ and not to exp $(- j\omega t)$, as is sometimes done. We shall continue to use these conventions.

From this it can be seen that ε is symmetrical:

$$\varepsilon_{kl} = \varepsilon_{lk}. \qquad (25)$$

Hence there are at most six different quantities in (24). The proofs that symmetries of the same nature must exist for elasticity and piezoelectricity run along the same lines.

The description of a crystal by means of six constants of the material ε_{kl} is satisfactory for static and for slowly varying fields. Losses and dispersion which become important at higher frequencies are not encompassed by this description. They can however be included in the formal framework of the permittivity, if the latter concept is extended in the following manner.

Complex and frequency-dependent permittivity

If, for an isotropic material, ε is regarded strictly as a constant of the material, then D is everywhere and at every instant (independently of the situation elsewhere and of previous events or states) given by the value of E at the same place and the same instant. In other words, $D = \varepsilon E$ is a local, instantaneous relation. In reality such a rigorous relation between D and E exists only in free space. For example, when E varies very rapidly, the polarization and therefore D in most materials also usually varies at the same frequency but the ratio of the amplitudes of D and E may well depend on the frequency and D often lags behind E. At a given moment D may thus depend not only on the value of E at that moment but also on previous values. Such a non-instantaneous relationship will be accounted for, as usual, by regarding the expression $D = \varepsilon E$ as a relationship between the complex quantities D and E, where ε is then a quantity that may be complex and frequency-dependent:

$$\varepsilon = \varepsilon'(\omega) - j\varepsilon''(\omega).$$

Since a minus sign is conventionally used here, a positive value of ε'' indicates a lag of D behind E and this implies losses, as can be seen by calculating the mean energy dissipated by the dielectric per second and per unit volume, $\overline{(\mathrm{Re}E)(\mathrm{Re}\dot{D})}$, which is found to be $\frac{1}{2}\omega\varepsilon''EE^*$.

At zero frequency we must recover the original relation between D and E: this means that $\varepsilon''(0)$ must be zero and $\varepsilon'(0)$ must be the permittivity for static fields.

The complex representation and method of calculation thus allows us to take account of non-instantaneous relations between D and E and so to describe losses ($\varepsilon'' \neq 0$) and dispersion (ε' is a function of ω). Non-local relations between D and E will not be considered here. Later on, the conductivity σ in (14) will sometimes also be taken as complex and frequency-depend-

ent. The permeability μ, on the other hand, will be considered here always as a real constant (and not complex as for example in problems related to electron and nuclear spin resonance).

Analogously we shall hereafter consider that ε in (23) may be a complex and frequency-dependent tensor. Instead of (25) we must then have the symmetry relation

$$\varepsilon_{kl} = \varepsilon_{lk}^* \qquad (26)$$

if the material is lossless. This follows since the time-average of $(\mathrm{Re}E) \cdot (\mathrm{Re}\dot{D})$ can be shown to be

$$\frac{1}{4}j\omega \sum_{kl}(\varepsilon_{kl} - \varepsilon_{lk}^*)E_kE_l^*,$$

and in a lossless crystal this must be zero for all E. At zero frequency the real part of ε must again reduce to the original tensor for static fields and the imaginary part must again vanish. The relation (26) then reduces again to (25).

The Onsager relations

The conclusion that the permittivity tensor ε must satisfy the relation (26) is based on the assumption that the medium is lossless. It can be shown in quite a different way that certain relations must in any case exist between the elements of ε, whether there are losses or not. However, other factors then have to be taken into account, e.g. whether or not the medium is subjected to a magnetic field H_0. If this is the case (and if this is the only other factor involved), then

$$\varepsilon_{kl}(-H_0) = \varepsilon_{lk}(H_0). \qquad (27)$$

These are the well known Onsager relations [12] applied to ε. Thus, if there is no magnetic field, ε is symmetric, whether there are losses or not. Only if ε_{kl} is real does (27) reduce to (26) when $H_0 = 0$.

The Onsager relations are applicable to the coefficients of many kinds of linear relationships in physics and engineering and are of a fundamental nature. They are based on the reversibility (in time) of microprocesses. They are valid only for coefficients relating variables that are conjugated in a prescribed manner. The derivation of the Onsager relations cannot be dealt with here.

Reversibility in a system of particles implies that all the particles would retrace their paths exactly if at a given moment all the velocities were reversed. All external influences that are antisymmetric in time must then also be reversed, for example electric currents and magnetic fields (which can always be considered as deriving from currents). The only influence of this kind mentioned in the foregoing was an applied magnetic field.

[12] See H. B. G. Casimir, Rev. mod. Phys. 17, 343, 1945.

Faraday rotation

On the basis of the symmetry relations we shall now set up a very simple ε-tensor with which the rotation of the plane of polarization in a constant magnetic field (Faraday rotation) can be formally described. The question of how the form of the tensor depends on the structure of the medium will not be discussed.

Restricting ourselves to lossless media, we resolve ε, element for element, into real and imaginary parts. From (26) the real part is symmetric, the imaginary part antisymmetric. We can therefore write:

$$\varepsilon = \varepsilon_s + j\varepsilon_a,$$

where ε_s is symmetric, ε_a is antisymmetric, and both are real.

Next, by a suitable rotation of coordinates, we reduce the ε-tensor to diagonal form. It can be shown that this is always possible for a real and symmetric matrix. ε_a then remains antisymmetric and real so that ε takes the form:

$$\varepsilon = \begin{pmatrix} \varepsilon_{s1} & 0 & 0 \\ 0 & \varepsilon_{s2} & 0 \\ 0 & 0 & \varepsilon_{s3} \end{pmatrix} + j\begin{pmatrix} 0 & \varepsilon_{a3} & \varepsilon_{a2} \\ -\varepsilon_{a3} & 0 & \varepsilon_{a1} \\ -\varepsilon_{a2} & -\varepsilon_{a1} & 0 \end{pmatrix}.$$

In our previously considered isotropic case (free space) all the ε_s's would be equal and all the ε_a's would be zero. One of the simplest deviations from isotropy — all the ε_a's zero but one of the ε_s's different from the other two — has also been discussed: this was the case of the uniaxial crystal (see 20) and it leads, as we have seen, to double refraction. If we now take all the ε_s's equal but make one of the ε_a's not zero:

$$\varepsilon = \begin{pmatrix} \varepsilon_s & j\varepsilon_a & 0 \\ -j\varepsilon_a & \varepsilon_s & 0 \\ 0 & 0 & \varepsilon_s \end{pmatrix}, \qquad (28)$$

we have the tensor with which the Faraday rotation can be described. We note, first of all, that according to the Onsager relations, ε_a in (28) can be non-zero only if a constant magnetic field H_0 is present: for we must have $\varepsilon_a(H_0) = -\varepsilon_a(-H_0)$. In the coordinate system x,y,z in which (28) is valid, the z-axis differs from the x- and y-axes: this must therefore be the direction of the magnetic field. The simplest case in which (27) is satisfied is that with ε_a proportional to H_z, so that its sign reverses if the field H_z is reversed. If the frequency goes to zero (25) must again be satisfied and so ε_a must become zero.

Next we show that (28) leads to a rotation of the plane of polarization. For a wave propagating along the z-axis, Maxwell's equations, with $B = \mu H$, $J = 0$, $\mu \neq 0$ and $\omega \neq 0$, lead again to the equations (21). The waves are thus purely transverse; from (21), D_z and H_z are zero and (28) then shows that this is also the case

for E_z. Combining (21) and (28) we find for the transverse components of E and H:

$$\left. \begin{aligned} \omega\varepsilon_s E_x - \ kH_y + j\omega\varepsilon_a E_y &= 0, \\ kE_x - \omega\mu H_y &= 0, \end{aligned} \right\} \text{(a)}$$

$$\left. \begin{aligned} -j\omega\varepsilon_a E_x \qquad\qquad + \omega\varepsilon_s E_y + \ kH_x &= 0, \\ kE_y + \omega\mu H_x &= 0. \end{aligned} \right\} \text{(b)} \qquad (29)$$

The terms are arranged in the same way as in (17a,b). The ε_a term now, however, couples the pair of equations (a) and (b), so that independent linearly polarized E_x,H_y-waves and E_y,H_x-waves are no longer possible. From (29) we find as dispersion relation:

$$\frac{1}{v^2} = \frac{k^2}{\omega^2} = \mu(\varepsilon_s \pm \varepsilon_a). \qquad (30)$$

We see here again what we already knew: propagation of undamped waves is possible only for real ε_s and real ε_a. Eliminating H_y from (29) and using (30) leads to:

$$E_x = \pm jE_y. \qquad (31)$$

We thus find a left-handed and a right-handed circularly polarized wave with different velocities. For a small difference in velocity ($|\varepsilon_a| \ll \varepsilon_s$), these waves, if of equal amplitude, can be combined to give a plane-polarized wave with a slowly rotating plane of polarization.

The plane of polarization forms a helix whose sense is the same as that of the circularly polarized wave with the smallest pitch, i.e. the slower of the two waves, that having the shorter wavelength. If the magnetic field has a polarity such that ε_a is positive, then the slower wave corresponds to the upper sign in (30) and in (31) and is thus left-handed if it travels in the $+z$-direction and right-handed if it travels in the $-z$-direction (optical convention, see (19) and the accompanying text); for negative ε_a (reversed polarity of magnetic field) the reverse is true. For a given polarity of the magnetic field, the sense of polarization of the slower wave, and consequently the *sense of rotation* of the *plane of polarization* (optical convention) is thus *opposite* for *waves propagated in opposite directions*; also, for both cases, the rotation reverses its sense when the magnetic field is reversed.

The situation is thus essentially different from that with natural optical activity, e.g. in quartz or in sugar solutions, where the sense of rotation (optical convention) is the same in both directions. The above description is therefore not applicable to natural optical activity. This follows, too, from the Onsager relation for the field-free case: $\varepsilon_{kl} = \varepsilon_{lk}$, which is inconsistent with (28).

Returning to the magnetic rotation, the situation is

drastically changed if $|\varepsilon_a|$ becomes larger than ε_s. From (30) it then follows that for real ω there is one real k and one imaginary k. Only one wave is therefore propagated; the other is cut off. Both are still circularly polarized. This is the situation that obtains for helicon waves as we shall see presently.

Electromagnetic waves in conducting media

Conductors obviously cannot support undamped waves; the field E and the consequent currents J give rise to losses. However, the losses may be small if J differs in phase from E by about $90°$. Some examples of both strongly attenuated waves and almost unattenuated waves in conductors will now be discussed.

We assume that the current is carried by free electrons in a crystal lattice, and also that the material has no pronounced magnetic or dielectric properties; for convenience we put $B = \mu_1 H$ and $D = \varepsilon_1 E$, and assume that the μ_1's and the ε_1's are little different from μ_0 and ε_0. The cardinal question is now: what is the relation between J and E? In 'normal' circumstances in conductors we have:

$$J = \sigma E, \qquad (32)$$

where σ is a constant of the material, the conductivity. Equation (32) can also be used under less normal circumstances, for example at very high frequencies, but then σ is a complex quantity, possibly frequency-dependent. Provided (32) is applicable, in one way or the other, a simple procedure enables us to make use of the results of previous calculations. Substitution of $\mu_1 H$, $\varepsilon_1 E$ and σE for B, D and J respectively in (7) and (8) yields:

$$\text{curl } H = (j\omega\varepsilon_1 + \sigma)E,$$
$$\text{curl } E = -j\omega\mu_1 H,$$

and these equations are equivalent to (16) if we replace ε in (16) by ε_{eff}, an *effective dielectric constant*:

$$\varepsilon_{\text{eff}} = \varepsilon_1 + \sigma/j\omega. \qquad (33)$$

We note here that ε_{eff} could assume the value zero if the two terms should compensate each other. In that case the dispersion relation $\omega\varepsilon = 0$ for *longitudinal electric waves* would be satisfied (see 17c); we shall see later that such waves are indeed possible. First, however, we shall give some examples in which ε_{eff} is still non-zero and the waves still transverse (17 a,b).

In metals, the term ε_1 in (33) can be neglected up to very high frequencies, so that ε_{eff} is given by

$$\varepsilon_{\text{eff}} = \sigma/j\omega. \qquad (34)$$

This can be seen from the fact that, while the permittivity of the material ε_1 is at most a few orders of

magnitude larger than ε_0 $(8.855 \times 10^{-12}$ F/m), the value of σ/ω at room temperature in copper (for example) has even at microwave frequencies a value of 10^{-2} F/m $(\sigma \approx 10^8$ Ω^{-1}m^{-1}, $\omega \approx 10^{10}$ s$^{-1})$.

The skin effect

Let us consider a metal of conductivity σ_0. Following the simple procedure mentioned above, we replace ε_0 in (18) by the ε_{eff} of (34), putting $\sigma = \sigma_0$, and we put $\mu_0 = \mu_1$. We then get the dispersion relation for transverse waves in a metal:

$$k^2/\omega^2 = \mu_1\sigma_0/j\omega \qquad (35)$$

or

$$k = \pm(1-j)\sqrt{\omega\mu_1\sigma_0/2}. \qquad (36)$$

For real ω, (36) represents strongly attenuated travelling waves of the type shown in fig. 3b; in particular the real and imaginary parts of k are equal in magnitude. This implies a wave of the type shown in *fig. 8*. Such waves can exist only in the neighbourhood of the *surface* of a metal. They propagate inwards from the surface and die out within a small distance, the *penetration depth* or *skin depth*. At high frequencies the skin depth is very small.

The above is a description of the *skin effect* at high frequencies (or in very thick wires). An a.c. current through a conducting wire is not distributed uniformly over the whole cross-section of the wire as is a direct current: the amplitude and phase of the current density are functions of distance from the surface and at high frequencies the current is confined to a thin layer under the surface. To discover the distribution of the current and its magnetic field (see inset, fig. 8) it is only necessary to consider a layer of thickness equal to a few times the penetration depth. If the penetration depth is small compared to the wire diameter, the surface can be considered as flat and in this case the current and field distribution can be calculated from (36), and the result is that shown in fig. 8.

For the values used above, $\sigma_0 = 10^8$ Ω^{-1}m^{-1}, $\omega = 10^{10}$ s^{-1}, and $\mu_1 = \mu_0 = 4\pi \times 10^{-7}$ H/m, the classical skin depth $\delta_k \equiv k_1^{-1} = \sqrt{2/\omega\mu_1\sigma_0}$ is only 1 μm.

The intrinsic impedance E_x/H_y of the metal for a transverse wave is found from (17a) and (36):

$$E_x/H_y = \omega\mu_1/k = \pm(1+j)\sqrt{\omega\mu_1/2\sigma_0}.$$

Because the value of σ_0/ω is so very much larger than ε_0, the modulus of E_x/H_y is many orders of magnitude less than $\sqrt{\mu_0/\varepsilon_0}$, the intrinsic impedance of free space. This implies a virtually complete mismatch between free space and metal. For this reason an electromagnetic wave in space incident on a metal surface is almost completely reflected (see p. 341/42).
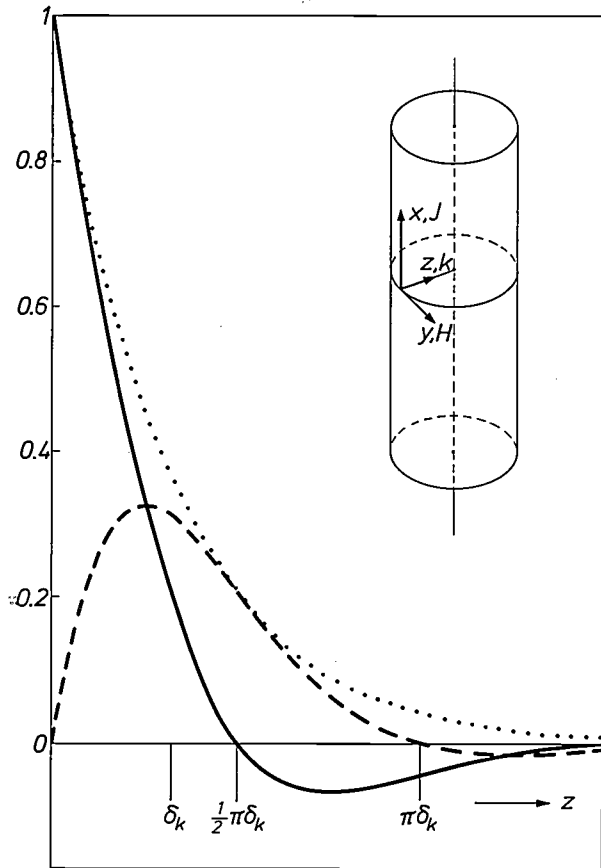
Fig. 8. Electromagnetic wave in a metal according to the dispersion relation (36) for real ω: the waveform at a given instant (solid curve), a quarter of a period later (dashed curve) and the amplitude of the wave as a function of z (dotted curve). The curves represent the functions $\exp(-\alpha) \cos \alpha$, $\exp(-\alpha) \sin \alpha$ and $\exp(-\alpha)$ respectively, where $\alpha = z/\delta_k$ and $\delta_k = \sqrt{(2/\omega\mu_1\sigma_0)}$ is the classical skin depth. The vertical scale represents a wave variable in arbitrary units, e.g. the current density J_x or the magnetic field H_y. (The field H_y has a quarter of a period phase lag with respect to J_x, as follows from (7) with $D = 0$ and $k = (1 - \mathrm{j})/\delta_k$.) The inset diagram gives the relative directions of J, H and k in a cylindrical wire carrying a high-frequency current. The plane-wave solution discussed is of course only valid here if δ_k is much smaller than the wire diameter.

It follows from this that J and E are no longer linked by a scalar relation such as (32), but by a tensor relation. *Table I* shows how in the normal situation, in the absence of a magnetic field, the usual scalar relation $J = \sigma_0 E$ is obtained. The equation (I.1) in the table expresses the fact that the conduction electrons (charge $-q$, mass m, concentration n) are accelerated by the field but also — because of collisions — are subject to an averaged frictional force; v_d is the resulting mean drift velocity of the electrons. The 'coefficient of friction' is the reciprocal of the relaxation time τ approximately equal to the mean time between collisions. When the left-hand side is neglected — justifiable if the frequency is not too high — then making use of (I.3), (I.4) and (I.6) we find the usual relation (I.5); μ_e is the 'mobility' of the electrons.

In order to include the effect of a static magnetic

Table I. Summary of the theory of conduction in metals (Drude) at low frequencies ('$\omega = 0$') and in the absence of a magnetic field. m mass of charge carriers, n their concentration, v_d drift velocity, τ relaxation time. The formulae are written for negative charge carriers (electrons) of charge $-q$. For positive charge carriers of charge $+q$, the signs in I.2 and I.4 and the sign of qE in I.1 would be reversed; with this convention, q, μ and σ_0 are thus always positive. For further explanation, see text.

$$m\dot{v}_\mathrm{d} = -qE - mv_\mathrm{d}/\tau \quad (\mathrm{I.1})$$
$$\omega = 0$$

$$v_\mathrm{d} = -\mu_\mathrm{e}E \quad (\mathrm{I.2})$$
$$\mu_\mathrm{e} = +q\tau/m \quad (\mathrm{I.3})$$

$$J = -nqv_\mathrm{d} \quad (\mathrm{I.4})$$

$$J = \sigma_0 E \quad (\mathrm{I.5})$$
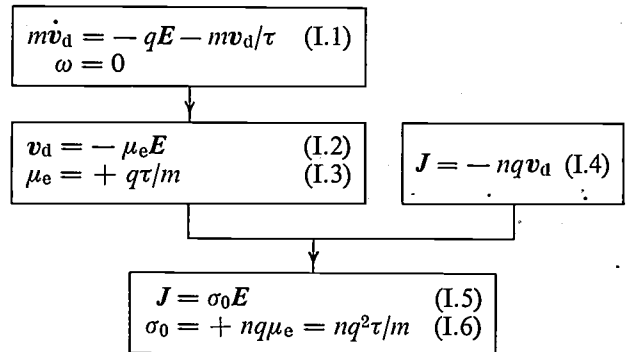$$\sigma_0 = +nq\mu_\mathrm{e} = nq^2\tau/m \quad (\mathrm{I.6})$$

Helicon waves

In a conductor with a high concentration of high-mobility electrons (e.g. a pure metal at low temperature) situated in a strong magnetic field, circularly polarized waves can propagate in the direction of the magnetic field. Under certain circumstances, these waves are practically unattenuated and propagate at an extremely low velocity. These waves are the helicon waves noted earlier. Their existence was predicted theoretically [13] in 1960 and in 1961 they were demonstrated experimentally [14]. There is a certain kinship with the Hall effect: as in that case, the current and electric field are not parallel — indeed, if the magnetic field and the mobility of the charge carriers are large enough, current and field may be almost perpendicular to one another.

field, a term representing the Lorentz force has to be added to the right-hand side of (I.1):

$$m\dot{v}_\mathrm{d} = -qE - qv_\mathrm{d}\times B_0 - mv_\mathrm{d}/\tau. \quad (37)$$

B_0 is the static magnetic flux density. We shall presently study waves that propagate in the direction of B_0, or in the opposite direction, and we therefore choose a coordinate system with the z-axis in this direction ($B_{0x} = B_{0y} = 0$, $B_{0z} = \pm B_0$). If the vector equation (37) is written out as three equations for the components of E and v_d, we find (because $B_{0x} = B_{0y} = 0$)

[13] O. V. Konstantinov and V. I. Perel', Sov. Phys. JETP 11, 117, 1960.
P. Aigrain, Proc. Int. Conf. on Semiconductor Physics, Prague 1960, p. 224.
[14] R. Bowers, C. Legendy and F. Rose, Phys. Rev. Letters 7, 339, 1961.

two equations in the transverse components v_{dx}, v_{dy}, E_x and E_y in which the magnetic field appears, and one equation in v_{dz} and E_z which is independent of the other two and in which the magnetic field does not occur. The equation in v_{dz} and E_z is of no interest to us and will not be treated further. Neglecting the left-hand side again and making use of (I.3) then for the transverse components, instead of (I.2) we find:

$$v_{dx} = -\mu_e E_x - \beta v_{dy},$$
$$v_{dy} = -\mu_e E_y + \beta v_{dx}, \tag{38}$$

where

$$\beta = \mu_e B_{0z}. \tag{39}$$

Equations (38) and (I.4) and Maxwell's equations lead to the wave phenomena called helicon waves. Before going further we should note that the quantity β — which for electrons is the opposite of the Hall ratio (see *fig. 9*) — is a kind of quality factor: only for $|\beta| \gg 1$ are the electric field and the drift velocity (i.e. current) nearly perpendicular to one another, the condition for virtually unattenuated helicon waves. It can be seen that this condition is only satisfied in quite extreme circumstances: for example, in a field $B_0 = 1\,\text{T} = 1\,\text{Vs/m}^2 = 10\,000$ gauss, we must have $\mu_e \gg 1\,\text{m}^2/\text{Vs}$, whereas in copper at room temperature μ_e is only about $6 \times 10^{-3}\,\text{m}^2/\text{Vs}$. Indeed, the first helicon experiment [14] was done with exceptionally pure sodium at 4 K: β was ≈ 40 for $B_0 = 1\text{T}$, so that $\mu_e \approx 40\,\text{m}^2/\text{Vs}$.

Solving (38) for v_{dx} and v_{dy} and using (I.4) yields a tensor relation between J and E instead of (I.5):

$$J = \sigma E,$$

where the tensor σ is given by:

$$\sigma = \frac{\sigma_0}{1 + \beta^2} \begin{pmatrix} 1 & -\beta \\ +\beta & 1 \end{pmatrix}; \tag{40}$$
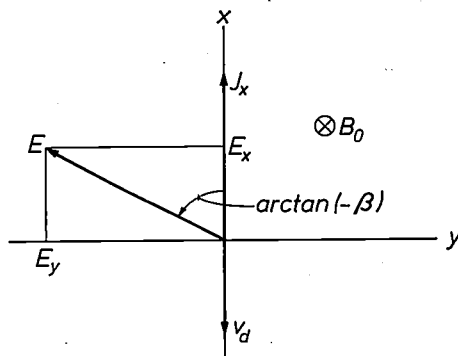
Fig. 9. Hall effect. The quantity β (eq. 39) for electrons is equal but opposite to the tangent of the Hall angle, i.e. the angle between the direction of the electric field and the current. This can be seen directly by putting $v_{dy} = 0$ in (38), i.e. by choosing the x-axis in the direction of the current. Equation (38) then gives $E_y/E_x = -\beta$. In the diagram B_0 is directed along the positive z-axis (into the paper); B_{0z} and β are thus positive, while E_y/E_x is negative.

σ_0 is the conductivity (I.6) when there is no static magnetic field.

The *dispersion relation for helicon waves* propagating along the z-axis can now be easily derived because of the following.

1) We consider good conductors for which we may write $\varepsilon_{\text{eff}} = \sigma/j\omega$.

2) The ε-tensor that then follows from (40):

$$\varepsilon_{\text{eff}} = \begin{pmatrix} \varepsilon_{\text{eff s}} & j\varepsilon_{\text{eff a}} \\ -j\varepsilon_{\text{eff a}} & \varepsilon_{\text{eff s}} \end{pmatrix} = \frac{\sigma_0}{j\omega(1 + \beta^2)} \begin{pmatrix} 1 & -\beta \\ +\beta & 1 \end{pmatrix}, \tag{41}$$

has the same form as the transverse part of (28), although $\varepsilon_{\text{eff s}} = \sigma_0/j\omega(1 + \beta^2)$ is no longer real for real ω (whereas $\varepsilon_{\text{eff a}} = \beta\sigma_0/\omega(1 + \beta^2)$ *is* still real). 3) The calculation based on (28) and using (29) which leads to (30) and (31) is straightforward and is therefore also applicable for ε_s and ε_a not real.

Application of (30), with $\mu = \mu_1$, therefore yields the required dispersion relation. The result is:

$$\omega = \frac{\pm \beta + j}{\mu_1 \sigma_0} k^2. \tag{42}$$

The tensor (41) satisfies the Onsager relations (β changes sign with B_{0z}) but no longer represents lossless propagation, as was to be expected, because $\varepsilon_{\text{eff s}}$ is no longer real. The medium is however *virtually* lossless when $|\beta| \gg 1$ because the real quantity $\varepsilon_{\text{eff a}}$ in (41) then dominates the imaginary quantity $\varepsilon_{\text{eff s}}$. In this situation ($|\beta| \gg 1$) the properties of helicon waves are most clearly manifested. The term j in (42) can then be neglected and we find, using (I.6) and (39):

$$\omega = \pm \frac{B_{0z} k^2}{\mu_1 n q}. \tag{43}$$

For real ω we find a real k, which means travelling waves, for the upper sign ($+$) if B_{0z} is positive. This corresponds to the upper sign in (31), i.e. to waves in which the vectors rotate clockwise as seen by an observer looking in the $+z$-direction. This is true for waves propagating in both directions along the z-axis ($k > 0$ and $k < 0$); see *fig. 10*. The sense of rotation using the optical convention (see p. 318) is thus, as in Faraday rotation, opposite for waves in the two directions, and reverses if the magnetic field is reversed. For positive B_{0z} the waves in which the vectors rotate anticlockwise (for an observer looking in the $+z$ direction) have imaginary k and are thus cut off. The whole of the discussion above has been based on the assumption that the charge carriers are electrons; if the conduction were to take place via holes, the senses of rotation would all be reversed.
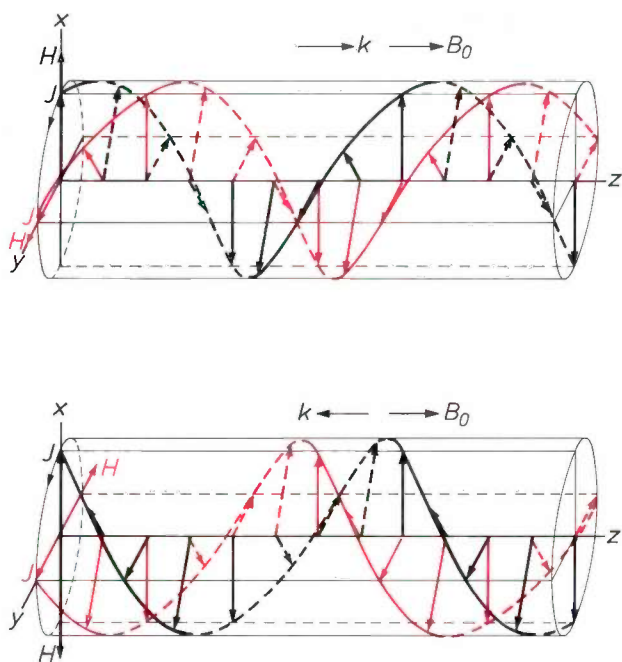
Fig. 10. Helicon waves moving in the direction of the magnetic field (upper diagram, k parallel to B_0), and opposite to the magnetic field (lower diagram, k in opposite sense to B_0). The black curves represent the current-density wave (J) at a given instant and the red curves the J wave a quarter period later. In the upper diagram, the wave moving to the right, H is in phase with J; in the lower diagram, the wave moving to the left, H is in antiphase with J. The electric field E is many orders of magnitude smaller than in free space for the same H (see eq. 44), and is of little importance. The diagrams refer to conduction with negative charge carriers (electrons); for hole conduction the vectors rotate in the opposite direction. Helicon vectors rotate in the same direction as the corresponding charge carriers in cyclotron resonance in the same magnetic field.

Helicon waves exhibit a strong dispersion: the phase velocity $v = \omega/k$, which from (43) is proportional to k or $\sqrt{\omega}$, can have widely differing values, depending on the frequency. In particular the velocity can be exceedingly low. For example, in a metal with $n = 6 \times 10^{28}$ m^{-3} in a field of 1 T (10 000 gauss), a helicon wave of frequency 17 Hz ($\omega = 100$ s^{-1}) has a wavelength of 6 mm and hence a velocity of 10 cm/s.

The term j in (42) represents the attenuation. From the form of (42) it can be seen, once more, that $|\beta|$ is a kind of quality factor. For a given medium (here this includes the value of B_0), $|\beta|$ is independent of ω. If the medium satisfies $|\beta| \gg 1$, the attenuation per wavelength (or per period) is just as small for low-frequency (slow) waves as for high-frequency fast waves.

At very high frequencies the left-hand side of (37) can no longer be neglected. It is found that the frequency at which this term begins to play a significant role lies in the neighbourhood of $\omega_c = qB_0/m$, the *cyclotron resonance frequency*. This is the angular frequency at which electrons in a magnetic field execute a circular or helical motion. For $\omega \ll \omega_c$, the helicon waves behave as described above.

The fact that σ_0/ω is so many orders of magnitude larger than ε_0 implies that the intrinsic impedance of a metal for helicon waves — as for the classical skin effect 'waves' — is many orders of magnitude less than that of free space for conventional electromagnetic waves. From (29a) and (42), neglecting losses:

$$E_x/H_y = \omega\mu_1/k = \sqrt{\beta\mu_1/(\sigma_0/\omega)} \ll \sqrt{\mu_0/\varepsilon_0}. \quad (44)$$

(Although $|\beta| \gg 1$, it is negligible compared to the very high value of the ratio of σ_0/ω to ε_0.) We therefore again have a complete mismatch between the medium and free space, so that both normal electromagnetic waves in free space and helicon waves in the medium are almost completely reflected at the interface.

As a result, in a configuration like that shown in *fig. 11*, standing waves can be set up whose attenuation is
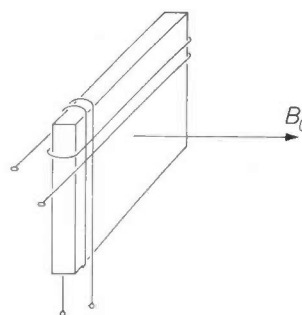
Fig. 11. Schematic diagram of an arrangement for a helicon experiment. A sample plate in a magnetic field perpendicular to the plate is provided with crossed coils. As a result of their circular polarization, standing helicon waves excited by one of the coils can be detected by the other coil. (In practice the primary coil is wound uniformly over the whole length of the plate.)

determined entirely by β. For example under the same conditions as above ($n = 6 \times 10^{28}$ m^{-3}, $B_0 = 1$ T) in a plate of thickness 3 mm ($= \lambda/2$), standing waves of 17 Hz can be expected.

In helicon experiments the sample is usually arranged with a primary coil and a secondary coil as in fig. 11. If a d.c. current is switched on or off in the primary, a series of standing helicon waves are excited and the corresponding damped oscillations induced in the secondary can be observed (*fig. 12*). Crossed coils are particularly well adapted for the experiment: the only coupling between them is via the (circularly polarized) helicon waves.

By means of such experiments, the elements of the σ-tensor and hence the *Hall constant* and the *magnetoresistance* [15] can be determined relatively easily and very accurately as functions of B_0 [16]. The determina-

[15] These are $\varrho_{xy}(B_0)/B_0$ and $\varrho_{xx}(B_0)$ respectively, where ϱ_{xx} and ϱ_{xy} are the elements of ϱ (the inverse of the tensor σ) which expresses E in terms of J through the relation $E = \varrho J$.

[16] R. G. Chambers and B. K. Jones, Proc. Roy. Soc. A **270**, 417, 1962.
M. T. Taylor, J. R. Merrill and R. Bowers, Phys. Rev. **129**, 2525, 1963.
See also E. Fawcett, Adv. Phys. **13**, 139, 1964.

tion of these quantities by conventional methods usually requires difficult precision measurements of very small resistances and voltages between accurately located contacts. Helicon measurements are made without contacts on the sample.
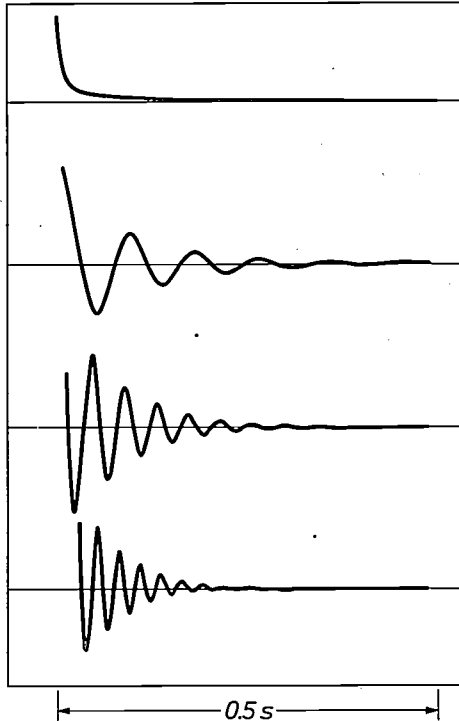

Fig. 12. Voltages induced by helicon waves, after R. Bowers, C. Legendy and F. Rose [14]. The diagrams show the voltages across the secondary coil (see fig. 11) as a function of time, after interruption of the primary current, with the sample in a magnetic field of strength (from top to bottom) 0 Oe, 3600 Oe, 7200 Oe and 10 800 Oe. (In this first helicon experiment, the sample was not of plate form as in fig. 11 but a cylinder of diameter 4 mm, and the coils were not crossed.)

Two complications that can arise with helicon waves should be mentioned. These do not come within the framework of purely *local* relations, characterized by effective ε's and ϱ's to which we have previously confined ourselves (see p. 320).

Firstly, there is the absorption arising from Doppler-shifted cyclotron resonance. The electrons responsible for conduction move in all directions through the metal at a high velocity, the Fermi velocity v_F (not to be confused with the drift velocity v_d). An electron with a Fermi velocity in the direction of the helicon wave runs through the wavefronts and is thus subject to an alternating field of frequency kv_F (the velocity of the slow helicon wave is neglected here). If kv_F is equal to ω_c, the electron undergoes cyclotron resonance and so absorbs energy from the wave and attenuates it. If kv_F is greater than ω_c then there are some electrons moving obliquely to the wave which come into resonance. For a given B_0, there is thus an absorption edge at $k = \omega_c/v_F$. Measurement of this absorption edge in single crystals for various direction of B_0 with respect to the crystal axes yields data on the anisotropy of the Fermi velocity and hence information about the shape of the Fermi surface [17].

Secondly, there is the question of 'open cyclotron orbits'. In a metal with a simple Fermi surface (e.g. an alkali metal) in a magnetic field, the momentum vector of an electron describes a closed orbit on the Fermi surface. In metals such as copper, silver and gold, however, the Fermi surface is so anisotropic that in certain directions the cyclotron orbits are 'open'. Because of this the helicon waves may be plane-polarized and strongly attenuated. This effect is also used for the study of the Fermi surface [18].

Reflection and transmission of optical waves in metals

We shall now leave situations involving magnetic fields to enquire what happens when the frequency of an electromagnetic wave is raised to the optical region. Important changes occur in the skin effect, primarily because the term $m\dot{v}_d$ in (I.1) can no longer be neglected. Instead of (I.5), with $\dot{v}_d = j\omega v_d$ we find:

$$J = \sigma_0 E/(1 + j\omega\tau).$$

Neglecting $m\dot{v}_d$ in (I.1) is clearly justified only when $\omega\tau \ll 1$. Let us now assume that the frequency is so high that $\omega\tau \gg 1$; there is then an effective conductivity

$$\sigma_{\text{eff}} = \sigma_0/j\omega\tau = -jnq^2/m\omega, \tag{45}$$

which is purely imaginary so that there are no losses (J and E differ in phase by 90°). Substituting (45) for σ_0 in (35) gives the dispersion relation:

$$k^2 = -\mu_1 nq^2/m. \tag{46}$$

Since k is imaginary, the waves are evanescent (fig. 3c, for real ω). As with the classical skin effect, these 'waves' are restricted to a thin layer at the surface of the metal. When electromagnetic waves are incident on such a surface, it follows that no power can be transmitted through the metal and there are also no losses; the waves are reflected completely. The shiny appearance of most metals is explained in this manner.

At still higher frequencies the term ε_1 in (33) can no longer be neglected. This means that we have an additional term $\varepsilon_1\mu_1\omega^2$ in (46):

$$k^2 = \varepsilon_1\mu_1\omega^2 - \mu_1 nq^2/m = \varepsilon_1\mu_1(\omega^2 - \omega_p^2),$$

where

$$\omega_p = \sqrt{nq^2/\varepsilon_1 m} \tag{47}$$

is called the *plasma frequency*. This is a critical frequency: for $\omega < \omega_p$, k is imaginary so that the waves are evanescent; for $\omega > \omega_p$, k is real so that the waves are propagated through the metal. In the first case there is complete reflection at the surface; in the second case there is partial reflection and partial transmission, dependent on the ratio of the intrinsic impedances of metal and free space (see Part III of this article). This impedance ratio passes through the value unity in the transition region, around the plasma frequency, and this implies zero reflection and 100% transmission.

For most metals ω_p lies in the ultraviolet. With $n = 6 \times 10^{28}$ m^{-3}, $\varepsilon_1 = \varepsilon_0$, and the usual values for the electronic charge q and mass m, we find $\omega_p \approx 1.4 \times 10^{16}$ s^{-1} which corresponds to a free-space wavelength of 140 nm. In this way the transparency of alkali metals in the ultraviolet region can be understood [19].

Longitudinal electric waves in conducting media

Introduction of an effective dielectric constant $\varepsilon_{eff} = \varepsilon_1 + \sigma/j\omega$ allowed us to make use of (16) for the problem of electromagnetic waves in conductors. For plane waves propagated in the z-direction we arrived at (17) and it was noted that longitudinal electric waves would be possible if ε_{eff} were to be zero. It can in fact be seen from (17) that if

$$\varepsilon_{eff} \equiv \varepsilon_1 + \sigma/j\omega = 0, \qquad (48)$$

then *all* components of E and H in (17) must be zero except E_z. (It follows directly from (8) that all magnetic components must be absent in a wave with a longitudinal electric field, since longitudinal vectors all have zero curl). Now ε_1 and σ themselves are not zero in (48), so that not only E but also D and J have longitudinal components. Since we also have $\partial_z \neq 0$ it follows that the divergences of D and J ($\partial_z D_z$ and $\partial_z J_z$), and hence ϱ_e and $\dot{\varrho}_e$ (see (10) and (11)), are neither of them zero. These waves are thus characterized by fluctuations in charge density: the electrons bunch together and disperse again. This is different from the case of purely transverse waves, where the charge density is everywhere zero (local electroneutrality): the divergence of a transverse vector is zero.

In what circumstances is the dispersion relation (48) satisfied? For real conductivity, $\sigma = \sigma_0$, we have for the first time the situation that ω *cannot* be real; ω must be purely imaginary, $\omega = j\sigma_0/\varepsilon_1$, and k is completely arbitrary. Any charge distribution $\varrho_e(z)$ with its corresponding field therefore dies away exponentially (see fig. 3g, h, j). The characteristic time for this process is $\tau_\varepsilon = \omega_1^{-1} = \varepsilon_1/\sigma_0$, the *dielectric relaxation time*. For metals τ_ε has no physical significance: $\varepsilon_1/\sigma_0 \approx 10^{-11}/10^8 = 10^{-19}$s and, for processes taking place in such a short time, the assumption that σ equals σ_0 is certainly incorrect. Certainly at radio frequencies we can conclude that there is always local *electroneutrality* in metals. In semiconductors, however, local space-charge variations do play a role and τ_ε is an important quantity as we shall see presently.

In media and under circumstances where $\omega\tau \gg 1$ (e.g. in metals at optical frequencies) the conductivity is purely imaginary, as we saw earlier (see 45). Longitudinal waves of real frequency are then possible, for substitution of (45) in (48) gives:

$$\varepsilon_1 - nq^2/m\omega^2 = 0,$$

so that

$$\omega = \omega_p = \sqrt{nq^2/\varepsilon_1 m}.$$

We see that the plasma frequency (47) is not only the critical frequency for the propagation of transverse waves but it is also *the* frequency at which longitudinal waves can exist, if $\omega\tau \gg 1$. As with dielectric relaxation, k is arbitrary. Such waves do not transfer energy: the Poynting vector $S = E \times H$ is zero because there are no magnetic fields.

The plasma frequency is a quantity continually encountered in 'plasma physics', which is the basic discipline for a number of quite diverse subjects such as travelling-wave amplifiers, astrophysics and controlled nuclear fusion. A *plasma* is a medium whose behaviour depends primarily on the charge and mass of the charge carriers and in which collisions play only a minor role. (Helicon waves are thus waves in a plasma.) The term was introduced in the twenties by Irving Langmuir, in connection with his investigations into gas discharges, to describe a dilute, strongly ionized but electrically neutral gas [20]. In this work Langmuir discovered that, surprisingly, electrons injected into the plasma rapidly came into thermal equilibrium with the plasma in spite of the very long mean free path. High frequency oscillations of the plasma would explain this. Such plasma oscillations had in fact been observed earlier by F. M. Penning [21]. The frequency found by Penning was 10^8 to 10^9 Hz, corresponding to wavelengths of several decimetres. From (47) this would imply an electron density of the order of 10^{17} m^{-3}, which is indeed typical for low-pressure gas discharges such as those used by Penning.

Summarizing we can say that local space-charge fluctuations die away exponentially in a time τ_ε if electron collisions play the dominant role ($\sigma = \sigma_0$), or oscillate at the frequency ω_p if collisions can be neglected ($\omega\tau \gg 1$). For charge variations that are *very steep* (large k) it is necessary to take into account a phenomenon that has not yet been discussed in this article, and which is of a non-electromagnetic nature: the *diffusion* of the electrons from regions of high concentration to regions of low concentration. We shall now look into this, but only for the case of low frequencies. The extra electron current due to diffusion is $-D_n$ grad n (D_n = diffusion constant). The concentration n has a gradient only because of deviations from the equilibrium concentration n_0 and the net local charge density corresponds exactly to these deviations $\varrho_e = -q(n - n_0)$, so that grad n = grad $(n - n_0)$ =

[17] E. A. Stern, Phys. Rev. Letters **10**, 91, 1963.
   M. T. Taylor, Phys. Rev. **137**, A 1145, 1965.
[18] S. J. Buchsbaum and P. A. Wolff, Phys. Rev. Letters **15**, 406, 1965.
   C. C. Grimes, G. Adams and P. H. Schmidt, Phys. Rev. Letters **15**, 409, 1965.
   See also the article by Fawcett [16].
[19] R. W. Wood, Phys. Rev. **44**, 353, 1933.
   C. Zener, Nature **132**, 968, 1933.
[20] I. Langmuir, Proc. Nat. Acad. Sci. **14**, 627, 1928.
[21] F. M. Penning, Nature **118**, 301, 1926, and Physica **6**, 241, 1926.

$= -q^{-1}\,\text{grad}\,\varrho_e$. The total electric current density therefore becomes:

$$J = \sigma_0 E - D_n\,\text{grad}\,\varrho_e. \qquad (49)$$

For plane longitudinal waves this can easily be reduced to the form (32). With $(\text{grad }\varrho_e)_z = -jk\varrho_e$ and $\varrho_e = \partial_z D_z = -jk\varepsilon_1 E_z$, we find:

$$J_z = (\sigma_0 + k^2\,\varepsilon_1 D_n)E_z.$$

The factor in the bracket is again an effective conductivity. Together with (48) it leads to the following improved dispersion relation for longitudinal waves:

$$j\omega\varepsilon_1 + \sigma_0 + k^2\varepsilon_1 D_n = 0. \qquad (50)$$

For waves of infinite wavelength ($k \to 0$), we find again the relaxation behaviour discussed above; from (50) we find — as was to be expected — that for shorter, steeper waves (steeper charge variations) the relaxation is more rapid.

We now consider infinitely *slow* waves ($\omega \to 0$) instead of infinitely *long* wavelengths. From (50) we find that these are *exponential charge distributions* having the characteristic length $k_1^{-1} = \sqrt{\varepsilon_1 D_n/\sigma_0} = = \sqrt{D_n\tau_\varepsilon}$. This is the Debye-Hückel length λ_D. A surface inside a conductor covered with a uniform charge is screened by a layer in which the charge density at the distance λ_D has fallen off by a factor e.

In the Debye and Hückel theory [22] of the conduction of electrolytes the quantities τ_ε and λ_D both play a role. Generally speaking, a positive ion is surrounded by a cloud of negative ions of radius λ_D; the positive ion experiences a frictional force, because relaxation causes the cloud to lag behind the positive ion when this moves.

In (50) we first neglected the third term and then we neglected the first term. Suppose we now neglect the second term ($\sigma_0 \to 0$): ε_1 then also disappears from the equation and all purely electric variables have vanished. With $\omega_1^{-1} = \tau_D$, and $k_1^{-1} = L_D$, (50) reduces to the familiar relation common to diffusion problems, $L_D = \sqrt{D\tau_D}$.

Elastic waves

Elastic waves in solids can be of a very complex nature. We shall introduce only a few elementary elastic waves here, but in passing we shall see how complications can easily arise. Later on we shall consider coupling between the waves introduced here and electromagnetic waves, and we shall then see that this can give rise to some remarkable effects in piezo-electrics.

We shall find in this section the well known result that the wave velocity (the velocity of sound) is highest in *rigid* and *light* substances. More specifically, in substances with a high resistance to *pressure* and *tension* but not to *shear*, longitudinal waves are fast but transverse waves are slow; in those with a high resistance to shear as well, transverse waves are also fast. In most substances the velocities of longitudinal and transverse waves do not differ greatly. Gelatine is an example in which transverse waves are much slower than longitudinal waves.

In elastic waves we are concerned with non-uniform *displacements* of volume elements, i.e. *deformation* of the material; this implies internal mechanical stresses in the material which, in turn, react on the displacements. The linear equations (algebraic and differential) between these quantities again define the wave problem.

Displacements, strains and stresses

Starting with the *displacement* u of each point of the material from its equilibrium position x,y,z, in which u is thus a function of x, y and z, the six strain components S_1, S_2, ... S_6 are defined as follows:

$$
\begin{aligned}
S_1 &= S_{xx} = \partial_x u_x, & S_4 &= S_{yz} = \partial_y u_z + \partial_z u_y, \\
S_2 &= S_{yy} = \partial_y u_y, & S_5 &= S_{zx} = \partial_z u_x + \partial_x u_z, \qquad (51) \\
S_3 &= S_{zz} = \partial_z u_z, & S_6 &= S_{xy} = \partial_x u_y + \partial_y u_x.
\end{aligned}
$$

There is deformation of the medium only if the displacement u is a function of the coordinates; if it is not, either there has been no displacement ($u = 0$) or the medium has been displaced as a whole ($u \neq 0$). The strain components are therefore derivatives of u. S_1, S_2 and S_3 give the extensions in the x-, y- and z-directions (*fig. 13a*) and S_4, S_5 and S_6 give the shear (fig. 13b); the latter consist of combinations of the derivatives such as $\partial_y u_z + \partial_z u_y$ because $\partial_y u_z \neq 0$ alone does not necessarily imply deformation as is explained in fig. 13.

The internal stress is the force per unit area exerted by material on one side of a given internal plane on material on the other side. In tension or compression the force is directed normally to the plane, in shear tangentially (*fig. 14*). We can therefore expect nine stress components T_{xx}, T_{xy}, ... T_{zz}; the first suffix indicates the direction of the force and the second the normal to the plane considered. The net couple on each volume element must be zero, which reduces the number of components to six (because $T_{yz} = T_{zy}$, $T_{zx} = T_{xz}$, $T_{xy} = T_{yx}$); this is explained in *fig. 15*. There remain the six stress components:

$$
\begin{aligned}
T_1 &= T_{xx}, & T_4 &= T_{yz} = T_{zy}, \\
T_2 &= T_{yy}, & T_5 &= T_{zx} = T_{xz}, \\
T_3 &= T_{zz}, & T_6 &= T_{xy} = T_{yx}.
\end{aligned}
$$

In equilibrium the net force on each volume element must also be zero which means that the stress field is

[22] P. Debye and E. Hückel, Phys. Z. 24, 305, 1923.
L. Onsager, Phys. Z. 27, 388, 1926.

homogeneous (*fig. 15*). We shall return to this later (p. 330).

Provided the strains are small, they are linearly related to the stresses (Hooke's law):

$$T_k = \sum_l c_{kl} S_l \qquad (k,l = 1, 2, \ldots 6). \qquad (52)$$

The 36 coefficients c_{kl} are called the elastic moduli (or stiffness constants). As in the derivation of (25) (and assuming that any changes in S and T are so slow that the c_{kl} remain real) it can be shown that the elastic moduli are symmetric ($c_{kl} = c_{lk}$) if no mechanical energy is transformed into other forms of energy. (The work done on the medium per unit volume in the elastic case is $\sum T_k dS_k$. This is explained in *fig. 16*). In the case of greatest anistropy (triclinic crystal) we still need however $6 + \frac{1}{2} \times (36 - 6) = 21$ different constants to describe the elastic properties of the ma-

terial. Clearly complications can arise all too easily. Crystal symmetries, however, restrict the number of independent constants and with a favourable choice of coordinate system this restriction becomes apparent through the appearance of many zeros and many equal-valued constants. In particular, *isotropic* material has only two independent constants; the array of constants then has the following form:

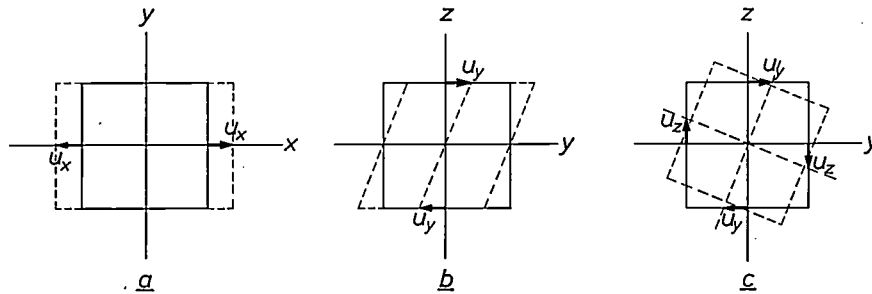|	S_1	S_2	S_3	S_4	S_5	S_6
T_1	c_{11}	c_{12}	c_{12}	0	0	0
T_2	c_{12}	c_{11}	c_{12}	0	0	0
T_3	c_{12}	c_{12}	c_{11}	0	0	0
T_4	0	0	0	c_{44}	0	0
T_5	0	0	0	0	c_{44}	0
T_6	0	0	0	0	0	c_{44}

(53a)

Fig. 13. Various types of deformation: *a*) extension, *b*) shear. In (*a*) a volume element is extended in the *x*-direction; $\partial_x u_x$ is positive. A negative $\partial_x u_x$ means a compression in the *x*-direction. By definition $\partial_x u_x$ is the first deformation component $S_{xx} = S_1$. In (*b*) layers perpendicular to the *z*-axis are displaced with respect to each other in the *y*-direction; $\partial_z u_y$ is positive. A positive or a negative value of $\partial_z u_y$ does not always imply deformation, however: in (*c*), where a volume element has been rotated without deformation, $\partial_z u_y$ is also positive. But $\partial_y u_z$ is equally large and of opposite sign. If the sum $\partial_y u_z + \partial_z u_y$ is not zero, the element does undergo deformation. This is the deformation component $S_{yz} = S_4$. Analogously, $S_2 = S_{yy}$, $S_3 = S_{zz}$ and $S_5 = S_{zx}$, $S_6 = S_{xy}$.
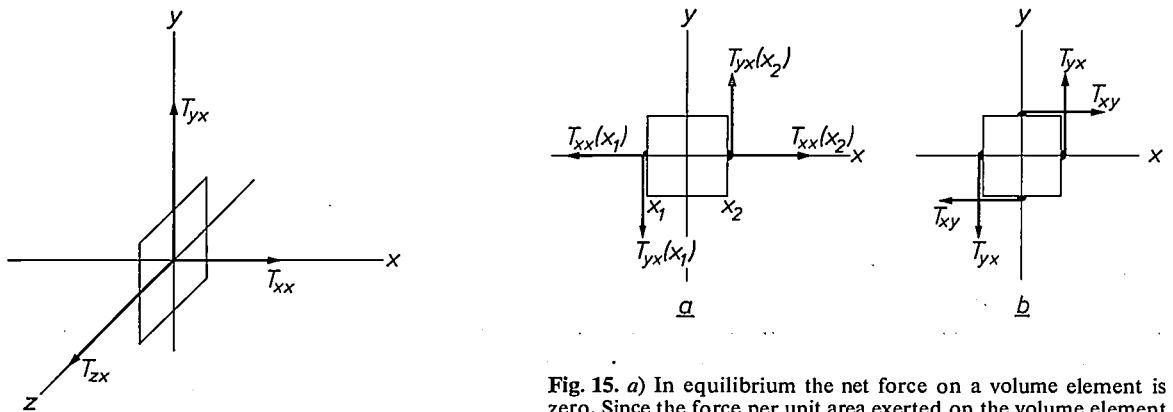

Fig. 14. Tensile stress and shear stress. The material on the $+x$ side of an elementary area perpendicular to the *x*-axis exercises a force on the material on the $-x$ side. The stress components T_{xx}, T_{yx} and T_{zx} are by definition the *x*-, *y*- and *z*-components respectively of that force per unit area of the elementary area. T_{xx} is a tensile stress ($T_{xx} > 0$) or a compression stress ($T_{xx} < 0$). T_{yx} and T_{zx} are shear stresses. The stress components T_{xy}, T_{yy}, T_{zy}, T_{xz}, T_{yz} and T_{zz} are defined analogously. See also fig. 15.

Fig. 15. *a*) In equilibrium the net force on a volume element is zero. Since the force per unit area exerted on the volume element on its face that faces left (or down or backwards) is equal but *opposite* to the stress there, we have $T_{xx}(x_2)$ is equal to $T_{xx}(x_1)$, $T_{yx}(x_2)$, to $T_{yx}(x_1)$, etc; in other words the stress field is uniform. *b*) The uniform stress field T_{yx} exerts a couple on the volume element. In equilibrium, this must be balanced by the opposite couple resulting from T_{xy}. It follows that $T_{xy} = T_{yx}$, and equally, $T_{yz} = T_{zy}$, $T_{zx} = T_{xz}$. There are thus six independent stress components: $T_1 = T_{xx}$, $T_2 = T_{yy}$, $T_3 = T_{zz}$, $T_4 = T_{yz}$, $T_5 = T_{zx}$, $T_6 = T_{xy}$.

where c_{11}, c_{12} and c_{44} are related as follows:

$$c_{11} = c_{12} + 2c_{44}. \qquad (53b)$$

One or two comments will serve to illustrate the significance of the constants c_{11}, c_{12} and c_{44} in the
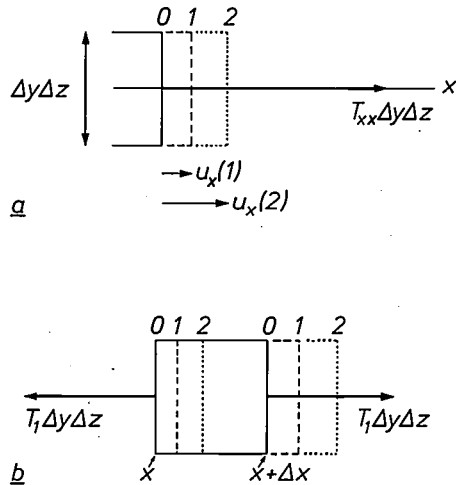
Fig. 16. Tensile stress and strain are assumed to be present here only in the x-direction. a) The work done on the material to the left of the element of area $\Delta y \Delta z$, when it is deformed from the state 1 (dashed) to the state 2 (dotted) is:

$$T_{xx}\Delta y \Delta z(u_x(2) - u_x(1)) = T_1 \mathrm{d}u_x \Delta y \Delta z.$$

Solid line: the undeformed state (0).
b) The work performed on the volume element $\Delta x \Delta y \Delta z$ is:

$$T_1(\mathrm{d}u_x(x + \Delta x) - \mathrm{d}u_x(x))\Delta y \Delta z =$$
$$= T_1 \, \mathrm{d}(u_x(x + \Delta x) - u_x(x))\Delta y \Delta z =$$
$$= T_1 \, \mathrm{d}(S_1 \Delta x)\Delta y \Delta z = T_1 \, \mathrm{d}S_1(\Delta x \Delta y \Delta z).$$

The work done per unit volume is thus $T_1 \mathrm{d}S_1$. For the general case it can be shown that the work per unit volume is $\Sigma T_k \mathrm{d}S_k$.

isotropic case. If c_{11} and $\dot{c}_{12}$ are non-zero and positive, a tensile stress T_1 implies not only an extension in the x-direction but also a *lateral contraction*: when $T_2 = T_3 = 0$, and S_1 is positive, it follows from (53a) that $S_2 = S_3 < 0$. That the shear modulus (or modulus of rigidity) c_{44} must be closely related to c_{11} and c_{12} is illustrated in *fig. 17*. It is shown there that the shear resulting from a shear stress (c_{44}) in one coordinate system is equivalent to an extension and a lateral contraction resulting from the combination of a tension and a lateral pressure (c_{11}, c_{12}) in an other coordinate system; and because of the isotropy, the constants must be independent of the coordinate system chosen. Finally, if c_{44} is zero (zero rigidity), $c_{11} = c_{12}$ so that $T_1 = T_2 = T_3 = c_{11}(S_1 + S_2 + S_3)$. It is easily shown that for any deformation, the relative change in volume $\Delta V/V$ is given by $S_1 + S_2 + S_3$. Thus $c_{44} = 0$ implies that when a deformation takes place involving no change of volume no stresses are set up. This is the situation with fluids (gases and liquids). In this sense gelatine and rubber are 'near-liquids' because c_{44} is small, and very much smaller than c_{11} and c_{12}.

Finally we must consider the dynamic influence of the stresses T on the displacements u. If the stress field is *non-uniform*, the volume elements undergo a net force which accelerates them. The net force in the x-direction on a volume element $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ is (see *fig. 18*): $(\partial_x T_{xx}\mathrm{d}x)\mathrm{d}y\mathrm{d}z + (\partial_y T_{xy}\mathrm{d}y)\mathrm{d}z\mathrm{d}x + (\partial_z T_{xz}\mathrm{d}z)\mathrm{d}x\mathrm{d}y$. This force is equal to the product of the mass $\varrho_m\mathrm{d}x\mathrm{d}y\mathrm{d}z$ and the acceleration $\partial_t^2 u_x$ in the x-direction (ϱ_m is the density). In this way we find the equations of motion:
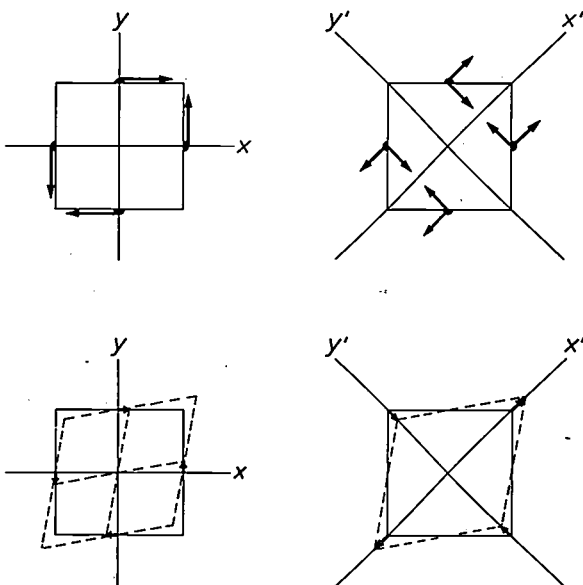
Fig. 17. A stress field (above) and the resulting strain (below) are described as shear stresses T_6 and shears S_6 in the x,y coordinate system. In the x',y' system the same stresses and strains are described as a tension T_1 and a compression T_2 resulting in an extension S_1 and a contraction S_2.
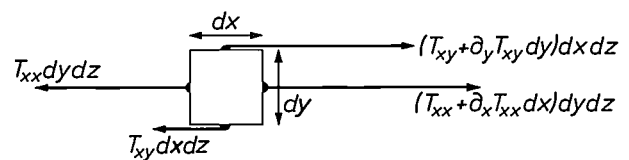
Fig. 18. Forces in the x-direction on a volume element in a non-uniform stress field. Besides the forces acting on the faces perpendicular to the x- and y-axes there are analogous forces on the z-faces.

$$\varrho_m \partial_t^2 u_k = \sum_l \partial_l T_{kl} \qquad (k,l = x, y, z). \qquad (54)$$

The six defining equations for the strain (51), the six Hooke equations (52) and the three equations of motion (54) give altogether 15 linear homogeneous equations for the 15 variables u_1, u_2, u_3, S_1, ... S_6, T_1, ... T_6, thus describing fully the wave phenomena. Of the possible solutions of these equations, we shall consider only two simple cases.

Transverse waves in isotropic materials

For transverse waves in an isotropic medium (see 53), propagating in the $+z$-direction ($\partial_x = \partial_y = 0$, $\partial_z = -jk$, $\partial_t = j\omega$) and with displacements only in the x-direction ($u_y = u_z = 0$), many of the 15 variables are zero. There remain:

$$(51) \to S_5 \qquad = -jku_x,$$
$$(52) \to T_5 \qquad = c_{44}S_5, \qquad (55)$$
$$(54) \to -\varrho_m\omega^2 u_x = -jkT_5,$$

from which the dispersion relation follows immediately:

$$\omega^2/k^2 = c_{44}/\varrho_m. \qquad (56)$$

The waves are thus dispersionless; at all frequencies the velocity of propagation is $v_{st} = \sqrt{c_{44}/\varrho_m}$.

Longitudinal waves in isotropic materials

For longitudinal waves in an isotropic medium propagating in the $+z$-direction ($u_x = u_y = 0, \partial_x = \partial_y = 0$, $\partial_z = -jk$, $\partial_t = j\omega$), (51), (52) and (54) reduce to:

$$(51) \to S_3 \qquad\quad = -jku_z,$$
$$(52) \to T_1 = \quad T_2 = c_{12}S_3, \; T_3 = c_{11}S_3, \qquad (57)$$
$$(54) \to -\varrho_m\omega^2 u_z = -jkT_3.$$

Combination of the three equations in u_z, S_3 and T_3 yields the dispersion relation:

$$\omega^2/k^2 = c_{11}/\varrho_m. \qquad (58)$$

(This is the condition for the existence of solutions with u_z, S_3 and T_3 non-zero; in each solution T_1 and T_2 follow directly from S_3.)

The velocity of longitudinal sound waves, $v_{sl} = \sqrt{c_{11}/\varrho_m}$, is therefore always higher than that of transverse sound waves, $v_{st} = \sqrt{c_{44}/\varrho_m}$. In solids v_{st} is often of the order of 2×10^3 m/s and, as c_{11} is about three times c_{44}, v_{sl} is about $\sqrt{3}$ times v_{st}. In lead v_{sl}/v_{st} is about $\sqrt{5}$, in 'near-liquid' materials such as gelatine and rubber v_{sl}/v_{st} is one or more orders of magnitude larger. In liquids there are no transverse waves, at least no transverse waves with a real wave vector.

Coupling of waves in an unbounded homogeneous medium

Up to now we have studied separately two classes of waves, electromagnetic and elastic waves. We shall now go on to consider *coupled* waves, in particular coupling between these two classes of waves, as found for example in piezoelectric materials. To give an illustration of what is involved in the coupling of waves, we shall take as example the longitudinal and transverse elastic waves in an isotropic medium which we have just treated separately. We shall now consider them together: we assume a wave propagated in the z-direction in which displacements are allowed both in the z- and the x-directions (but not in the y-direction). We then find again equations (55) and (57), now together. These are written symbolically as follows:

u_x	S_5	T_5	u_z	S_3	T_3	
×	×					
	×	×		•		} (55)
×		×				
			×	×		
	•			×	×	} (57)
			×		×	

Each row represents an equation and the crosses in-

dicate which variables are involved. (The equations for T_1 and T_2, here omitted, are of no importance at the moment.) The 6×6 determinant of the equations is the product of two 3×3 determinants, $f(\omega,k)$ and $g(\omega,k)$. The dispersion relation is thus

$$f(\omega,k)g(\omega,k) = 0. \qquad (60)$$

There are therefore two independent solutions, representing two types of wave:

$f = 0$, variables u_x, S_5, T_5 ($u_z = S_3 = T_3 = 0$): transverse waves;

$g = 0$, variables u_z, S_3, T_3 ($u_x = S_5 = T_5 = 0$): longitudinal waves.

Let us now suppose that the medium loses its isotropy in such a way that a tension T_3 results not only in an extension (S_3) but also in a shear (S_5). The purely longitudinal wave can then no longer exist: the tensile stress associated with it would cause transverse displacements u_x via the term S_5. This is expressed in (59) by terms at the black dots (c_{35} and thus c_{53} no longer zero). The 6×6 determinant then no longer factorizes: longitudinal and transverse waves are *coupled*.

In this way the E_x,H_y waves and the E_y,H_x waves of (17a,b) are mixed in (29) by the ε_a term. Similarly, electromagnetic and elastic waves in piezoelectric materials are coupled because the electric fields give rise to mechanical stresses.

Weak coupling

In the above we found pure longitudinal waves and pure transverse waves for $c_{35} = 0$. If c_{35} is not zero but still very small — weak coupling — we may expect 'nearly-pure' longitudinal and transverse waves. In such a case we can start from the dispersion relations $f = 0$, $g = 0$ as a zero-order approximation to find the relation between ω and k for the new waves. If, for example, the terms at the dots in (59) are small, although not zero, the dispersion relation becomes, instead of (60):

$$f(\omega,k)g(\omega,k) = \delta(\omega,k), \qquad (61)$$

where δ is a measure of the coupling. More complicated situations can arise giving, for example, a dispersion relation of the form $fg^2 = \delta$; however, the simpler form (61) is usually found and we shall restrict our discussion to this form.

For a given frequency ω_0, the new waves will have wave numbers in the neighbourhood of those of the old f-wave and g-wave ($fig.\ 19$). Suppose that the wave number of the old f-wave is k_0, so that $f(\omega_0,k_0) = 0$, and let the new wave number closest to k_0 be $k_0 + \Delta k$. For $\omega = \omega_0$, it follows from (61) that

$$(f_0 + f'_{k0}\Delta k)(g_0 + g'_{k0}\Delta k) = \delta, \qquad (62)$$

where f_0, g_0, f'_{k0}, g'_{k0} are the values of f, g, $\partial f/\partial k$, $\partial g/\partial k$ at ω_0, k_0. According to our assumption, $f_0 = 0$; on the other hand, g_0 will in general differ substantially from zero, so that $g'_{k0}\Delta k$ can be neglected. To a first approximation, therefore,

$$\Delta k = \delta/f'_{k0}g_0 \qquad (\omega = \omega_0). \qquad (63)$$

Similarly, for a given wave number k_0, the difference in frequency $\Delta\omega$ between the new 'near-f-wave' and the old f-wave is

$$\Delta\omega = \delta/f'_{\omega0}g_0 \qquad (k = k_0). \qquad (64)$$

In a similar way we can find the difference in wave number or the difference in frequency between the new 'near-g-wave' and the old g-wave. These expressions will be useful later on.

Resonant coupling

Suppose that the curves $f = 0$, $g = 0$ intersect at some point with real k and ω. Such an intersection can occur, apart from the trivial case of $k = 0$, $\omega = 0$, only if at least one of the waves is dispersive. An example is the combination helicon wave/sound wave; see $fig.\ 20$. At the point of intersection k_0,ω_0 the two waves are 'resonant': their phase relation is constant in both space and time. Even a weak coupling can then give a strong effect. In particular, for a given δ, k and ω will exhibit larger changes than in the non-resonant

case. This can be seen directly by applying (62) to the point of intersection. Because $f_0 = 0$, $g_0 = 0$ we find (see $fig.\ 21$):

$$\Delta k = \pm \sqrt{\delta/f'_{k0}g'_{k0}} \quad \text{at } \omega = \omega_0, \qquad (65)$$

and

$$\Delta\omega = \pm \sqrt{\delta/f'_{\omega0}g'_{\omega0}} \quad \text{at } k = k_0. \qquad (66)$$
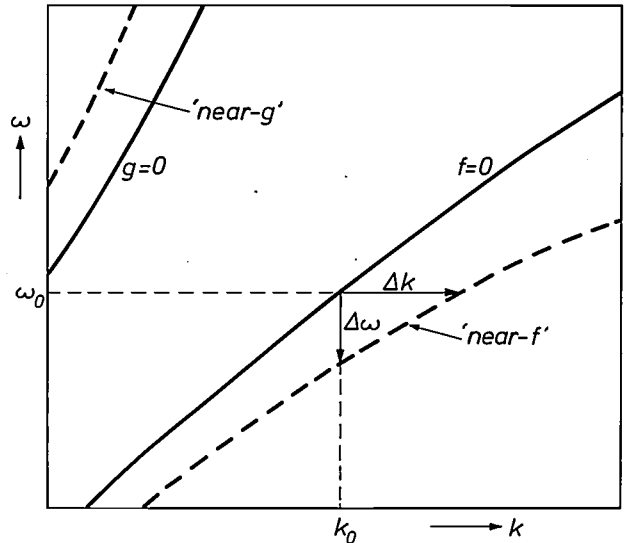

Fig. 19. The dispersion relations $f = 0$ and $g = 0$ (solid curves) for two independent types of waves. The dashed curves represent the dispersion relations of waves that can result from the coupling of the f and g waves.
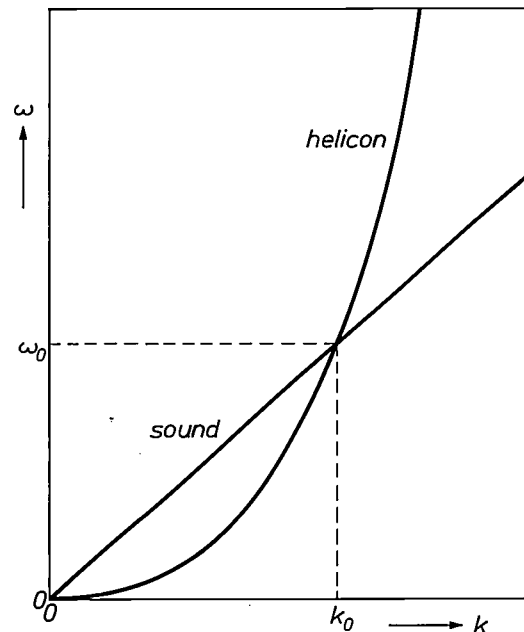

Fig. 20. The dispersion relation for helicon waves (43) is a square law in k, that for acoustic waves is linear. They therefore intersect not only at the origin but also at another point (k_0, ω_0). At this intersection even a weak interaction gives rise to strong effects (resonance). For a given magnetic field, the intersection takes place at lower frequencies as the concentration of the electrons is lower (see eq. 43). For $B_0 = 1$ T, the resonant frequency in metals lies in the gigacycle region. In semiconductors the resonance lies in the megacycle region or lower.

As δ approaches zero, $\sqrt{\delta}$ approaches zero less rapidly and, at sufficiently small δ, the resonant effects (65) and (66) are much larger than the non-resonant effects (63) and (64). Owing to the coupling, the intersection disappears; in its neighbourhood the curves $f = 0$, $g = 0$ become two hyperbolae asymptotic to the original intersecting lines. If, in (65) and (66), Δk, $\Delta \omega$ or both happen to be imaginary, the curves shown respectively
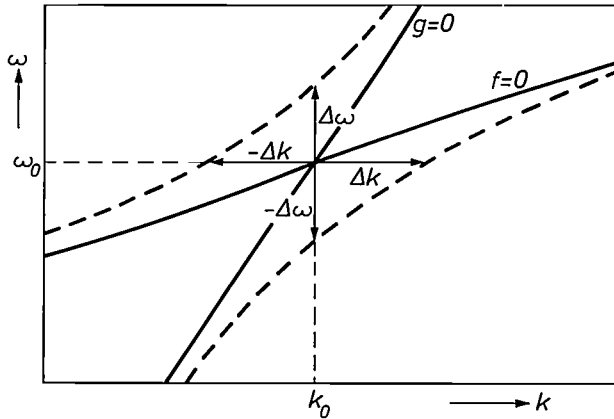

Fig. 21. The effect of weak coupling on the dispersion curves of two waves in the neighbourhood of resonance.
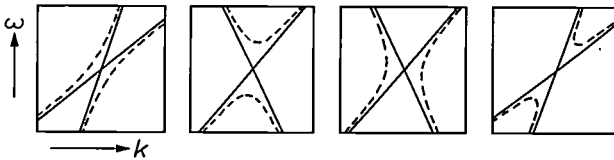
Fig. 22. The various situations that can occur with resonant coupling, depending on whether Δk or $\Delta \omega$ in (65) and (66) are real or imaginary.
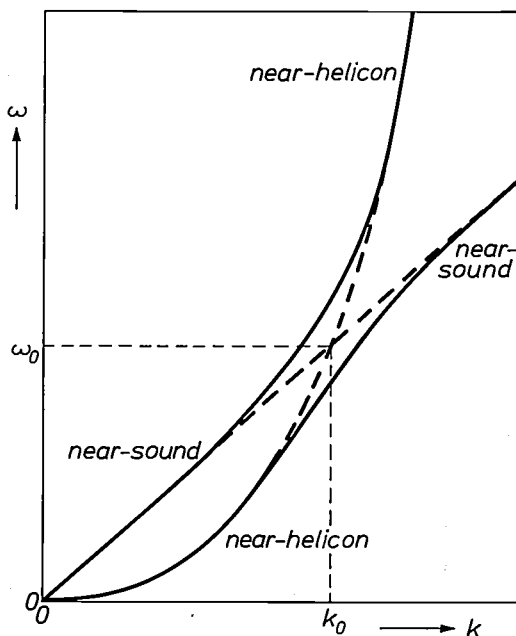

Fig. 23. The effect of coupling between helicon waves and sound on the combined dispersion relation.

in the second, third, and fourth diagrams of *fig. 22* are relevant. One of the main results of the analysis given in [7] is, that in the fourth situation we have travelling-wave amplification while the third gives rise to the 'absolute instability' which was briefly mentioned in the introduction. The second situation was already referred to as 'cut-off' on page 314.

For the combination helicon wave/sound wave there is a weak electric coupling between the electron motion of the helicon wave and the ion motion of the sound wave [23]. The combined dispersion relation of fig. 20 thus splits into two branches which nowhere intersect except at the origin (*fig. 23*); along each branch, the corresponding wave gradually changes from 'near-sound' to 'near-helicon wave' or vice versa.

The coupling of helicon waves with sound was first demonstrated experimentally in potassium [24]. However, the point of intersection of the uncoupled dispersion curves for metals in 'convenient' magnetic fields (up to 1 T = 10 kilogauss) lies in the none too easily accessible gigacycle region. In this respect, semiconductors are more convenient: owing to the lower electron concentration the helicon curve in fig. 20 is steeper so that the point of intersection is displaced to lower frequencies. W. Schilz [25] has demonstrated the coupling in N-PbTe with $n \approx 10^{24}$ m^{-3}. He used a plate of PbTe fitted with two quartz transducers (*fig. 24*). The three elements were all acoustically a half wavelength long at the same frequency (≈ 800 kHz) and thus resonant. The acoustic transmission of the system peaked sharply at this frequency (curve a, fig. 24). The PbTe plate was also fitted with crossed coils (as in fig. 11) by means of which a $\lambda/2$ helicon resonance could be set up and detected at this same frequency when the magnetic field was about 0.4 T (= 4 kG). This field and this frequency correspond to the point of intersection shown in fig. 20. If, at this field, a signal of varying frequency was applied to one of the helicon coils, then the transducers gave a sharply peaked output close to the resonant frequency (curve b, fig. 24). The height of this peak varied with the magnetic field in approximately the same way as the helicon resonance itself. The high electron mobility necessary for this experiment was obtained by using a single crystal of PbTe at a temperature of 4 K ($\mu \approx 100$ m^2/Vs).

[23] See for example D. N. Langenberg and J. Bok, Phys. Rev. Letters **11**, 549, 1963, and J. J. Quinn and S. Rodriguez, Phys. Rev. Letters **11**, 552, 1963 and Phys. Rev. **133**, A 1589, 1964.

[24] C. C. Grimes and S. J. Buchsbaum, Phys. Rev. Letters **12**, 357, 1964.

[25] W. Schilz, Phys. Rev. Letters **20**, 104, 1968.

Fig. 24. Demonstration of the coupling be-
tween sound waves and helicon waves in a
semiconductor (lead telluride) by W.
Schilz [25]. The PbTe plate P is wound with
crossed coils C (see also fig. 11) for the
generation and detection of helicon waves,
and fitted with electro-acoustic transducers
for the generation and detection of elastic
waves. The plate and the transducers form
a single acoustic unit with a certain resonant
frequency ($\approx$ 800 kHz). The magnetic field
is adjusted to a value such that helicon
resonance occurs at this same frequency.
The output signal A of one of the electro-
acoustic transducers is shown on the right
as a function of the frequency f, a) when
the other transducer is excited, b) when
one of the helicon coils is excited.
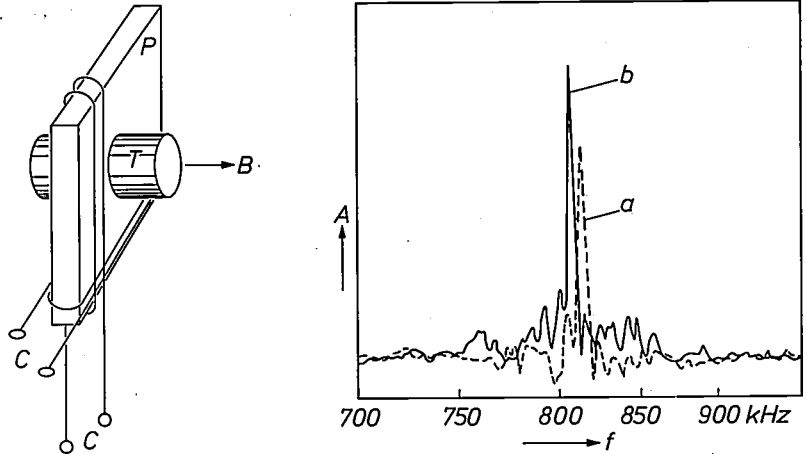
Piezoelectric coupling

In piezoelectric materials mechanical variables such as stress and displacement are coupled with electrical variables such as field and polarization. Since the twenties piezoelectric materials have been used extensively for the conversion of acoustic or mechanical signals into electrical signals and vice versa: in microphones, loudspeakers, pick-ups, crystal oscillators, filters, transducers, etc. For a long time the well known materials tourmaline, Rochelle salt and quartz have been of practical importance and indeed quartz still is. Since the fifties, however, a large number of synthetic piezoelectric materials have entered the field.

Materials exhibiting piezoelectricity all have a crystal structure without a centre of symmetry. It can easily be seen that a centro-symmetric crystal cannot be piezoelectric: for instance, a uniform deformation cannot cause any separation of the centres of positive and negative charges, i.e. a polarization, in such a crystal.

Conduction can mask the piezoelectricity of a crystal, for the motion of free charges tends to destroy an electric field if there is sufficient time available. All the piezoelectric materials widely used in practice are consequently insulators. Nevertheless, it is in fact the interaction of a piezoelectric crystal lattice with free charge carriers — in *piezoelectric semiconductors* — that can offer so many interesting effects. Such possibilities form the main theme of the next section. Cadmium sulphide and tellurium are examples of such piezoelectric semiconductors.

The piezoelectric coupling is expressed through electric terms in the elastic equations (52) and mechanical terms in the electric equations (22):

$$T_k = \sum_l c_{kl}S_l - \sum_n e_{nk}E_n, \qquad (67a)$$

$$D_m = \sum_l e_{ml}S_l + \sum_n \varepsilon_{mn}E_n. \qquad (67b)$$

$$(k,l = 1, \ldots 6; \quad m,n = x,y,z)$$

The nine linear equations (67) with their 81 coefficients express the nine variables $T_1, \ldots T_6$, D_x, D_y, D_z in terms of the nine variables $S_1, \ldots S_6$, E_x, E_y, E_z. Apart from the minus signs in (67a), the 9×9 array of coefficients is again symmetric if the electrical and mechanical energy is not transformed into other forms such as heat. With regard to the coefficients e, this symmetry has already been incorporated in the notation of (67), whereas for the c's and ε's we must have, as before:

$$c_{kl} = c_{lk}, \quad \varepsilon_{mn} = \varepsilon_{nm}. \qquad (68)$$

This can be shown by reasoning analogous to that on p. 319/320. Again we assume that the coefficients are real constants; this means that the variables may not vary too rapidly.

If we assume that the electromechanical state of the material is completely determined by the six quantities T or S and the three quantities D or E, the expression

$$dU = \Sigma T_k dS_k + \Sigma E_m dD_m,$$

i.e. the mechanical and the electrical work done when the state of the material changes must again be a total differential; in other words the 'internal energy' U is again a variable of state. It follows in the same way as before, that the 9×9 array of coefficients will be completely symmetric if T and E are expressed in terms of S and D. In (67), however, as in the literature on acoustic amplification and electro-acoustic surface waves, we have expressed T and D in terms of S and E. A difference in sign for the e then occurs between (67a) and (67b); this can be seen as follows. Under the above assumption, not only U but also the quantity $H_E = U - \Sigma E_m D_m$ is a variable of state (i.e. it is fully determined by the nine quantities T or S and D or E), so that

$$dH_E = \Sigma T_k dS_k - \Sigma D_m dE_m$$

is a total differential. From this follows, in the usual manner, the symmetry of the array of coefficients in (67), including the minus signs. H_E is called the *electric enthalpy*.

Fortunately, a drastically simplified form of (67) is sufficient for the discussion of a number of typical cases. This is because, in certain materials possessing a sufficient symmetry, waves in certain directions are

possible in which only one component of each of T, S, u, D, and E is of importance. We shall restrict the discussion to such situations. We then have:

$$T = cS - eE, \qquad (69a)$$
$$D = eS + \varepsilon E. \qquad (69b)$$

Strictly speaking, in each actual case we should indicate which component of T, S, D and E is involved. We shall do this for D and E — we shall discuss waves for which D and E are either transverse or longitudinal — but we shall not specify which components of T, S and u are involved.

The positive dimensionless constant $e^2/\varepsilon c$ will be found to be a convenient measure of the relative strength of the piezoelectric coupling. In most piezoelectric materials, $e^2/\varepsilon c$ is much less than 1, i.e. the coupling is weak. In certain specially prepared ceramics, e.g. certain lead zirconium titanates, $e^2/\varepsilon c$ may approach the value unity.

As for the purely electromagnetic waves, we shall first discuss insulators here and then examine what complications and possibilities arise as a result of conduction. These considerations follow, somewhat freely, the analysis of A. R. Hutson and D. L. White [26].

Near-light and near-sound

We shall first consider *electric transverse* waves in weakly piezoelectric insulators. As might be expected, these are near-electromagnetic waves or near-acoustic waves. The deviations in velocity from 'pure light' and 'pure sound' respectively will be calculated from the weak-coupling relation (63).

We take the direction of propagation as the z-axis and the direction of D and E as the x-axis. Maxwell's equations reduce to the first two, (7) and (8), with $J = 0$. Because ∂_x and ∂_y are zero and, according to our assumptions, only one component of each of S, T and u are involved, only one of the defining equations (51) remains ($S = \partial_z u$). Similarly, only one of the equations of motion (54) remains ($\varrho_m \partial_t^2 u = \partial_z T$). Using (69) as well we thus obtain six equations with six variables. With $\partial_x = 0$, $\partial_y = 0$, $\partial_z = -jk$ and $\partial_t = j\omega$, these equations become:

$$kH_y = \omega D_x, \quad \omega\mu H_y = kE_x; \quad \text{(Maxwell)} \quad \text{(a,b)}$$
$$D_x = eS + \varepsilon E_x, \quad T = cS - eE_x; \quad \text{(mixed) (c,d)} \quad (70)$$
$$-jku = S, \qquad \varrho_m\omega^2 u = jkT; \quad \text{(elastic)} \qquad \text{(e,f)}$$

Equations (a,b) are purely electromagnetic, (c) is electric with some elastic admixture, (d) is elastic with some electrical admixture and (e,f) are purely elastic. We can find the dispersion equation by putting

[26] A. R. Hutson and D. L. White, J. appl. Phys. 33, 40, 1962.
    D. L. White, J. appl. Phys. 33, 2547, 1962.

the 6×6 determinant of (70a . . . f) equal to zero or by successive elimination of the wave variables. The result can be written as:

$$(\omega^2 - v_l^2 k^2)(\omega^2 - v_s^2 k^2) = \frac{e^2}{\varepsilon c} \omega^2 k^2 v_s^2. \quad (71)$$

In this equation v_l and v_s are the velocities of light and ound respectively in the 'unperturbed' medium ($v_l = 1/\sqrt{\varepsilon\mu}$, $v_s = \sqrt{c/\varrho_m}$). As anticipated, equation (71) is of the form (61), $fg = \delta$. For zero coupling ($e = 0$) we return to the two independent waves: light waves of velocity v_l and sound waves of velocity v_s. With (weak) coupling ($e^2/\varepsilon c \ll 1$) we find near-light and near-sound. To find the velocity of the near-light wave we put $\omega_0 = v_l k_0$. Using (63), together with

$$f = \omega^2 - v_l^2 k^2, \quad g = \omega^2 - v_s^2 k^2, \quad \delta = \frac{e^2}{\varepsilon c} \omega^2 k^2 v_s^2,$$

we obtain:

$$\frac{\Delta k}{k_0} = -\frac{1}{2} \frac{e^2}{\varepsilon c} \frac{v_s^2}{v_l^2}. \quad (72)$$

For the near-sound wave, putting $\omega_0 = v_s k_0$, we find analogously:

$$\frac{\Delta k}{k_0} = +\frac{1}{2} \frac{e^2}{\varepsilon c} \frac{v_s^2}{v_l^2}. \quad (73)$$

(In (72) and (73), v_s^2 in the denominator has been neglected in comparison with v_l^2.) Since $\Delta v/v$ is equal to $-\Delta k/k$, the velocity of the near-light wave is a fraction $\frac{1}{2}(e^2/\varepsilon c)v_s^2/v_l^2$ greater than that of light in a non-piezoelectric medium with the same ε and μ. Similarly, the velocity of the near-sound wave is the same fraction smaller than the velocity of sound in a non-piezoelectric medium of the same ϱ_m and c. The effect is extremely small ($\ll 10^{-10}$), not primarily because the coupling factor $e^2/\varepsilon c$ is small but because the velocities of light and sound differ by so large a factor (10^5). This is a matter of some generality: between waves of greatly differing velocities there is little interaction.

Piezoelectric stiffening

We now consider *electric longitudinal* waves in piezoelectric insulators. Since in *non*-piezoelectric materials, as in free space, electromagnetic waves of this type do not exist, there is no question of near-light in this case; we can expect only an acoustic wave with some admixture of an electric-longitudinal character. Its velocity is easily found. Since E is longitudinal, so that curl E is zero, it follows from the Maxwell equation (8) (with $\mu \neq 0$, $\omega \neq 0$), that $H = 0$; also from (7) (with $J = 0$, $\omega \neq 0$), that $D = 0$. (For longitudinal D, this also follows from the fact that

div $D = 0$ for an insulator.) From the mixed equations

$$0 = eS + \varepsilon E,$$
$$T = cS - eE, \qquad (74)$$

it follows by elimination of E that

$$T = c(1 + e^2/\varepsilon c) S. \qquad (75)$$

We see that only elastic wave variables remain for which the conventional elastic equations are valid — with this difference that c is replaced by $c(1 + e^2/\varepsilon c)$. The material has thus become effectively stiffer. This phenomenon [27] is known as *piezoelectric stiffening*. The velocity of the near-sound wave is thus equal to $\sqrt{c(1 + e^2/\varepsilon c)/\varrho_m} \approx (1 + \frac{1}{2}e^2/\varepsilon c)\sqrt{c/\varrho_m}$, so that the relative effect on the velocity is a factor $v_1{}^2/v_s{}^2$ ($\approx 10^{10}$) larger than above.

For the elastic wave, the distinction between electric-longitudinal and electric-transverse admixture is essentially the distinction between electrostatic and electrodynamic phenomena. This explains the relatively small effect of the electric-transverse admixture: electrodynamic effects are small for slow processes and an elastic wave *is* slow compared with a light wave. Let us now elaborate on this a little further. In the neutral insulator with which we are concerned, the divergence of D must be zero. For the longitudinal wave, this also implies that $D = 0$ and that therefore the polarization wave eS gives rise to a field wave E of significant amplitude: $\varepsilon E = -eS$. According to (75) this results in a considerable change in the elastic stiffness of the medium. In the transverse case there is no *electrostatic* reason for D to be zero: all transverse vectors have zero divergence. The (transverse) polarization eS now excites a field E only by *induction*. This however is very small. From (70a,b) we find (with $\omega/k = v_s$): $\varepsilon E_x = \varepsilon \mu v_s{}^2 D_x = (v_s{}^2/v_1{}^2)D_x$. The terms in E_x in (70c,d) can therefore be neglected so that D_x is virtually equal to eS and, from (70d), the stiffness is hardly affected. Summarizing, the electric field is significant in the longitudinal case ($E = -eS/\varepsilon$) but not in the transverse case. For this reason we may restrict ourselves to longitudinal fields and currents in our discussion of piezoelectric semiconductors, which now follows.

Acoustic waves in piezoelectric semiconductors

Mobile charge carriers in a piezoelectric material cause attenuation of an acoustic wave: the electric fields associated with the wave cause currents and thus ohmic losses. On the other hand, however, there is the interesting possibility of negative attenuation, i.e. amplification of an acoustic wave, through activation of the medium by a constant electric drift field [28]. A simple criterion decides whether amplification is possible or not: the drift velocity v_{d0} must be higher than the wave velocity v_s. As noted above, we are exclusively concerned here with longitudinal fields and currents.

We shall treat the attenuation and amplification of acoustic waves in two stages. First we shall establish formally the relation between attenuation or amplification and the complex conductivity $\sigma = \sigma_r + j\sigma_i$, which is the ratio of the wave variables J and E. It will be found that a *negative* σ_r implies amplification. After this we shall show that with a constant applied electric field, a negative σ_r can be obtained.

For the first stage we write, as usual (eq. 32):

$$J = \sigma E. \qquad (76)$$

As in the case of piezoelectric stiffening, we shall derive a new *effective* stiffness of the material by elimination of the electric wave variables. From Maxwell's equations (using curl $E = 0$ for longitudinal E, whence $H = 0$ and therefore $J + \dot{D} = 0$), and (76):

$$D = -\sigma E/j\omega.$$

Substituting this in (69) now leads not to (74) and (75) but to:

$$T = cS - eE, \qquad (77a)$$
$$0 = eS + (\varepsilon + \sigma/j\omega)E, \qquad (77b)$$

and hence to:

$$T = c\left(1 + \frac{e^2}{\varepsilon c} \frac{1}{1 + \sigma/j\omega\varepsilon}\right) S.$$

We now have an effective elastic stiffness of

$$c' = c\left(1 + \frac{e^2}{\varepsilon c} \frac{1}{1 + \sigma/j\omega\varepsilon}\right).$$

There is thus stiffening of the medium again but it is now a complex stiffening. (Comparing this with the stiffening in insulators, we see that, as before, the conduction could have been taken into account by introducing an effective dielectric constant $\varepsilon + \sigma/j\omega$.)

We assume that $\Delta c = c' - c$ is small compared with c. This is certainly the case for weak piezoelectric coupling, provided $|1 + \sigma/j\omega\varepsilon|$ is not too much less than unity. We then obtain directly from the dispersion relation $k^2/\omega^2 = \varrho_m/c$ for sound waves, the relative change $\Delta k/k$ occurring at a given ω as a result of the relative complex stiffening $\Delta c/c$:

$$\frac{\Delta k}{k} = -\frac{1}{2}\frac{\Delta c}{c} = -\frac{1}{2}\frac{e^2}{\varepsilon c}\frac{1}{1 + \sigma/j\omega\varepsilon}.$$

We are interested only in the imaginary part Δk_i of Δk. Negative $\Delta k_i/k$ implies attenuation, positive $\Delta k_i/k$ amplification of the sound wave. (See (6) and the caption to fig. 3. If Δk_i is sufficiently small, the relative

change in amplitude *per unit length* is equal to Δk_i and the relative change *per wavelength* is $2\pi\Delta k_i/k$.) Remembering that σ can be complex, we find:

$$\frac{\Delta k_i}{k} = -\frac{1}{2}\frac{e^2}{\varepsilon c}\frac{\sigma_r/\omega\varepsilon}{(1 + \sigma_i/\omega\varepsilon)^2 + (\sigma_r/\omega\varepsilon)^2}. \quad (78)$$

This shows that amplification occurs if σ_r is negative.

The small change in the real part of k, i.e. in the wave velocity, can be neglected in the calculation of the attenuation or amplification. In the following, therefore, we put $\omega/k = v_s$.

In the simple cases where σ merely represents the static conductivity ($\sigma_r = \sigma_0$, $\sigma_i = 0$), the attenuation is given by

$$\frac{\Delta k_i}{k} = -\frac{1}{2}\frac{e^2}{\varepsilon c}\frac{\omega_\varepsilon/\omega}{1 + \omega_\varepsilon^2/\omega^2},$$

where ω_ε is called the relaxation frequency, $\omega_\varepsilon = \sigma_0/\varepsilon$. This is the reciprocal of the dielectric relaxation time τ_ε discussed on p. 327. The attenuation is a maximum for $\omega = \omega_\varepsilon$, and has a significant value only when ω and ω_ε are of the same order of magnitude. It can be shown from this that, even for frequencies in the GHz region, attenuation of this nature is found only in semiconductors with a very low concentration of charge carriers. In photoconductors σ_0, and hence ω_ε, is dependent on the intensity of the incident light. The light-sensitive acoustic attenuation found in CdS [29] can be explained in this way [30].

Amplification of acoustic waves

We now come to the second stage: to show that σ_r can become negative by the application of an electric field E_0 that gives the electrons a constant drift velocity v_{d0}, thus enabling acoustic waves to be amplified.

The acoustic wave in a piezoelectric semiconductor produces variations in three directly related quantities via the lattice polarization. These are the longitudinal electric field, the charge density (electron concentration) and the current. There are now *two* factors in the expression for the current density that vary: the drift velocity *and* the electron concentration. We are thus concerned with the product of two varying quantities and we can therefore no longer use complex quantities. We now denote the *actual* field by $E_0 + E_1$, the *actual* drift velocity by $v_{d0} + v_{d1}$ and the *actual* electron concentration by $n_0 + n_1$, where E_0, v_{d0} and n_0 are constant and E_1, v_{d1} and n_1 are real quantities with frequency ω and wave number k. We then have $v_{d1} = -\mu_e E_1$. The current density is

$$-q(n_0 + n_1)(v_{d0} + v_{d1}) =$$
$$= -qn_0v_{d0} - qn_0v_{d1} - qn_1v_{d0} - qn_1v_{d1}.$$

There is thus a d.c. component $-qn_0v_{d0}$ and a wave

component $-qn_0v_{d1} - qn_1v_{d0}$ of frequency ω and wave number k. The last term $-qn_1v_{d1}$ is of second order and can be neglected. This term does contain a d.c. component $-q\langle n_1v_{d1}\rangle$ to which we shall return in the next section. At the moment, however, we are only concerned with the wave component. Using complex wave variables J, n and E again ($n_1 = \text{Re } n$, $E_1 = \text{Re } E$) we find, with $n_0q\mu_e = \sigma_0$ (see Table I, p. 323, eq. I.6) and $-qn = \varrho_e$:

$$J = \sigma_0 E + v_{d0}\varrho_e. \quad (79)$$

The current in the wave thus consists of a direct reaction $\sigma_0 E$ on the field and a current $v_{d0}\varrho_e$ which is *automatically* present in each charge fluctuation ϱ_e because the electron gas as a whole has the drift velocity v_{d0}.

Equation (79) is not quite complete. Because there are charge fluctuations there is also diffusion, so that besides the current (79) there is also a diffusion current. We shall neglect this diffusion current for the moment, but will return to it presently.

Apart from (79) there is a further relation between the wave variables J and ϱ_e: the continuity equation (11). Because J is longitudinal, this assumes the form $-jkJ = -j\omega\varrho_e$ or, with $\omega/k = v_s$:

$$J = v_s\varrho_e. \quad (80)$$

Elimination of ϱ_e from (79) and (80) gives the ratio between J and E and hence the complex conductivity that we require. The result is:

$$J = \sigma_0 E/(1 - v_{d0}/v_s),$$

and thus

$$\sigma_r = \sigma_0/(1 - v_{d0}/v_s), \qquad \sigma_i = 0. \quad (81)$$

This shows that by giving the electrons a drift velocity v_{d0} greater than the wave velocity v_s, σ_r can indeed be made negative and amplification of the wave can be achieved.

The operation of the acoustic amplifier is therefore as follows. Let us assume that the acoustic wave propagates at a velocity v_s to the right. The associated longitudinal alternating electric field causes charge fluctuations that propagate with the wave. Now a positive net space charge only moves to the right because it is continually built up at the front and continually dispersed at the back by an *a.c. current* moving

[27] See for example W. Voigt, Lehrbuch der Kristallphysik, Teubner, Leipzig 1910.

[28] A. R. Hutson, J. H. McFee and D. L. White, Phys. Rev. Letters 7, 237, 1961.

[29] H. Gobrecht and A. Bartschat, Z. Physik 153, 529, 1959.
H. D. Nine, Phys. Rev. Letters 4, 359, 1960.
H. D. Nine and R. Truell, Phys. Rev. 123, 799, 1961.
H. J. McSkimin, T. B. Bateman and A. R. Hutson, J. Acoust. Soc. Amer. 33, 856, 1961.

[30] A. R. Hutson, Phys. Rev. Letters 4, 505, 1960. See also the article by Hutson and White [26].

in the direction of the wave. This current must thus be directed to the right in a positive space charge and to the left in a negative space charge (see *fig. 25*). This is expressed by the continuity equation (80) for longitudinal waves.

If the mean drift velocity is zero, this current is directly caused by the alternating field, which thus maintains the charge wave. Alternating field and current are in phase, so that there are losses. When, however, the electron gas as a whole has a velocity v_{d0} to the right, each net positive space charge now implies a current to the right (and each negative space charge a current to the left) so that a weaker alternating field is necessary to maintain the charge wave. Alternatively we see that, with the same alternating field, larger charge fluctuations will arise. If v_{d0} is equal to v_s, then no field at all is necessary to maintain the charge wave — it propagates 'by itself' — and, for a given alternating field, infinitely large space charges would arise (in the absence of diffusion). Finally, if v_{d0} is larger than v_s, then a positive space charge must now be dispersed at the front and built up at the back to maintain the stationary state, and this has to be done by the *alternating field* (*fig. 26*). Field and current are therefore in antiphase, so that the field, and hence the wave, take up energy from the charge carriers.

By ignoring the diffusion we have been able to give a fairly simple description of the essentials of the acoustic amplifier. To discuss the practical possibilities, however, it is necessary to take the diffusion into account. This can be done by adding a diffusion term $-D_n \, \mathrm{grad} \, \varrho_e = jkD_n\varrho_e$ to the right-hand side of (79), see also (49). It is then easy to show that (81) becomes:

$$\frac{\sigma_r}{\sigma_0} = \frac{\gamma}{\gamma^2 + \omega^2/\omega_D^2}, \quad \frac{\sigma_i}{\sigma_0} = \frac{\omega/\omega_D}{\gamma^2+\omega^2/\omega_D^2}, \quad (82)$$

where $\gamma = 1 - v_{d0}/v_s$ and $\omega_D = v_s^2/D_n$.

To find the amplification given by the negative conductivity, we must combine (82) with (78). The result for the amplification per unit length is found, after some manipulation, to take the form:

$$\Delta k_i = -\tfrac{1}{2}\frac{e^2}{\varepsilon c}\frac{\omega_\varepsilon}{v_s}\frac{\gamma}{\gamma^2 + (\omega_\varepsilon/\omega + \omega/\omega_D)^2}. \quad (83)$$

In (83) ω/k is again written as v_s. From this equation it can be seen that two restrictions are imposed on the frequency: there is substantially no amplification when ω is much smaller than ω_ε or much greater than ω_D. The physical reason is as follows. When ω is much smaller than ω_ε, the charge fluctuations relax so rapidly that they completely screen off the piezoelectric field. If, on the other hand, ω is much larger than ω_D, then the waves are so short and the wave flanks so steep that diffusion erases the charge fluctuations. Ampli-

fication is thus possible only in materials for which $\omega_\varepsilon < \omega_D$. Now in most materials ω_D is no larger than 10 GHz. With $\varepsilon \approx \varepsilon_0 \approx 10^{-11}$ F/m, it then follows that to make an acoustic amplifier, the piezoelectric semiconductor must have a conductivity $\sigma_0 = \varepsilon\omega_\varepsilon$ smaller than $10^{-11}\times10^{10} = 10^{-1}\ \Omega^{-1}\ \mathrm{m}^{-1}$.

The acoustic amplifier is an example of the general phenomenon in which a stream of particles can interact in some way with a travelling wave, and give up energy to that wave when their velocity is larger than the wave velocity. Another well known example is the travelling-wave tube, in which electrons are injected into an electromagnetic wave of relatively low velocity.

Acoustic amplification was first demonstrated experimentally by A. R. Hutson, J. H. McFee and D. L. White [28] in a photoconducting CdS crystal. The acoustic transmission they measured is plotted as a function of the applied field in *fig. 27*. Free charge carriers were produced by illumination of the crystal. In
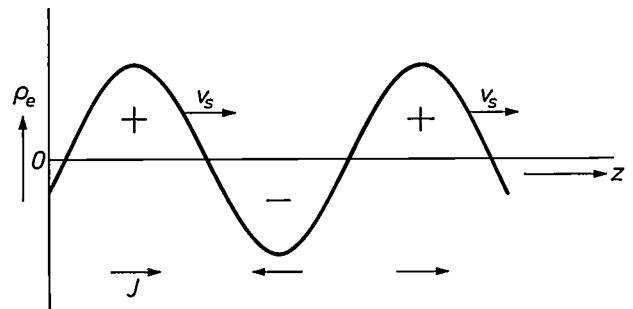
Fig. 25. Illustration of the continuity equation for clouds of space charge moving from left to right. Clouds of positive space charge move to the right because on their right there is a net accumulation and on their left a net removal of positive charge. There must therefore be an a.c. current directed towards the front and away from the back of a moving cloud of positive charge, i.e. to the right in the clouds of positive charge and to the left in the clouds of negative charge, in accordance with (80).
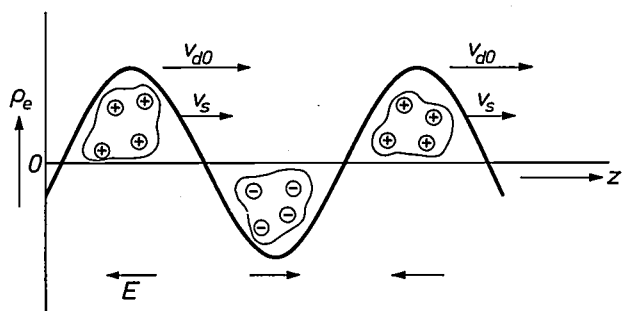
Fig. 26. Clouds of space charge consisting of surpluses and deficiencies in the concentration of an electron gas that moves as a whole with a drift velocity v_{d0} that is larger than the wave velocity v_s. These clouds would move faster than the wave ($v_{d0} > v_s$) were it not for the alternating field set up by the wave. In order to get clouds of space charge moving with the required velocity v_s each cloud must be dispersed at the front and concentrated at the rear. This means that in clouds of positive charge the field must be directed to the left, and in the clouds of negative charge it must be to the right. The positive clouds, however, still represent current to the right and the negative clouds a current to the left (fig. 25). Field and current are therefore in antiphase.

the dark the crystal was a good insulator; the concentration of charge carriers, and hence ω_ε, was adjusted by the incident light intensity. The zero level in fig. 27 refers to acoustic transmission through the crystal in the dark (the attenuation at 45 MHz was less than 7 dB and at 15 MHz less than 2 dB); positive dBs mean attenuation, negative dBs amplification. The results are qualitatively in good agreement with equation (83). In particular, the charge mobility can be derived from the zero intersection at ≈ 700 V/cm, by putting the drift velocity corresponding to this field equal to the velocity of the transverse acoustic wave used, $v_s = 2 \times 10^5$ cm/s. The mobility of 285 cm²/Vs found in this way agrees well with values found by other method (≈ 300 cm²/Vs).

The acoustic amplifier we have dealt with can be regarded as the precursor of the amplifier of acoustic *surface waves* (p. 347-348) which has attracted attention in recent years.
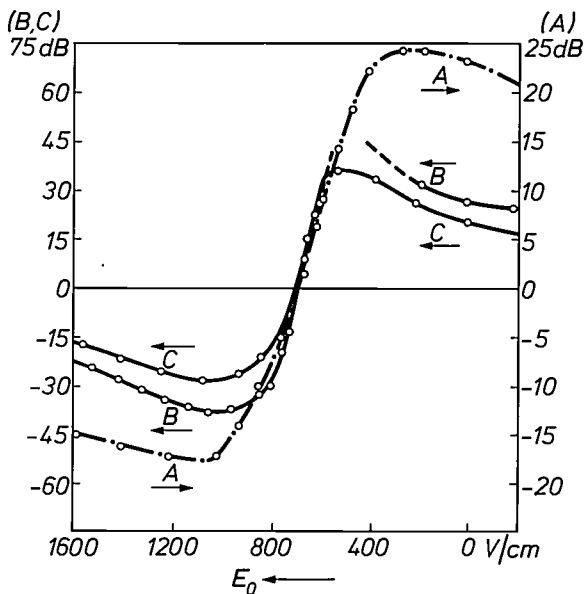
Fig. 27. Acoustic amplification as shown by A. R. Hutson, J. H. McFee and D. L. White [28]. The vertical axis represents the acoustic attenuation in dB in a photoconducting CdS crystal of length 7 mm as a function of the applied drift field E_0 for various frequencies and various levels of illumination. The conductivity σ_0 and hence the relaxation frequency $\omega_\varepsilon = \sigma_0/\varepsilon$ were adjusted by means of the illumination of the crystal. Curve A, 15 MHz, $\omega_\varepsilon/\omega = 1.2$; B, 45 MHz, $\omega_\varepsilon/\omega = 0.24$; C, 45 MHz, $\omega_\varepsilon/\omega = 0.21$. The attenuation is given relative to that of the crystal in the dark (0 dB). At 45 MHz the 'dark' attenuation was less than 7 dB, at 15 MHz less than 2 dB. Positive dBs mean attenuation, negative dBs mean amplification.

Acoustic amplification considered as weak coupling of two types of wave

We have considered the acoustic attenuation and amplification as a result of the 'complex stiffening' of the material. In this connection we came across the phenomena of dielectric relaxation and diffusion; these are 'longitudinal-electrical' wave phenomena (p. 327). Acoustic attenuation and amplification can also be

considered as the result of weak piezoelectric coupling between an acoustic wave and a longitudinal electrical 'wave' [31]. To show this, we shall start with a complete set of equations in the variables J, D, E (all longitudinal), ϱ_e, T, S and u in, for example, the following form:

$$J + \dot{D} = 0, \quad T = cS - eE, \quad S = \partial_z u,$$
$$\text{div } J = -\dot{\varrho}_e, \quad D = eS + \varepsilon E, \quad \partial_z T = \varrho_m \partial_t^2 u,$$
$$J = \sigma_0 E + v_{d0}\varrho_e - D_n\text{grad } \varrho_e.$$

The dispersion relation that follows from these equations can easily be written in the form $fg = \delta$ (eq. 61), remembering that $e^2/\varepsilon c$ is small. The result is:

$$\underbrace{(\omega^2/v_s^2 - k^2)}_{f} \underbrace{[\omega_\varepsilon + j(\omega - v_{d0}k) + k^2 D_n]}_{g} =$$
$$= \underbrace{(e^2/\varepsilon c)k^2 [j(\omega - v_{d0}k) + k^2 D_n]}_{\delta}. \qquad (84)$$

By putting $f = 0$, we have once more the unperturbed sound wave, while $g = 0$ yields an unperturbed longitudinal electric wave — another version of (50) somewhat complicated by the occurrence of v_{d0}. Applying (63), with $\omega_0 = v_s k_0$, to (84) we obtain after some amnipulation a complex expression for $\Delta k/k$, the imaginary part of which is indeed given by (83).

The acousto-electric effect; Weinreich's relation

We have seen that a d.c. current in a piezoelectric semiconductor can affect an acoustic wave in that material. Conversely, an acoustic wave in the material can give rise to a d.c. current or to a static electric field (depending on whether the semiconductor is short-circuited or open-circuited). The charge carriers are 'carried along' by the wave. This is known as the acousto-electric effect. It also occurs, to a less extent, in other materials. When R.H. Parmenter predicted the effect [32] in 1952, he remarked on the correspondence with two other quite different phenomena — the linear accelerator, in which electrons are accelerated by oscillating fields that are synchronized to form an (accelerating) travelling wave, and the thermoelectric effect where charge carriers are borne along by the stream of phonons (acoustic quanta) caused by a temperature gradient.

The effect can be attributed to the d.c. component that is present in the current density because both the electron concentration and the drift velocity are modulated when the material carries an acoustic wave. On p. 337 we saw that this component is equal to $-q \langle n_1 v_{d1} \rangle$, where n_1 and v_{d1} are actual physical modulations; using complex wave variables n and v_d, the d.c. component is $-\frac{1}{2}q\text{Re}(nv_d^*)$. We thus have an acousto-electric d.c. current source

$$J_{ae} = -\tfrac{1}{2}q\text{Re}(nv_d^*) = \tfrac{1}{2}\text{Re}(\varrho_e v_d^*).$$

This is equivalent to a 'field source'

$$E_{ae} = J_{ae}/\sigma_0 = \tfrac{1}{2}\text{Re}(\varrho_e v_d^*)/\sigma_0. \qquad (85)$$

[31] See for example K. Blötekjær and C. F. Quate, Proc. IEEE 52, 360, 1964.

[32] R. H. Parmenter, Phys. Rev. 89, 990, 1953.

This is the field that is equivalent to the acoustic wave in its effect on the charge carriers. This acousto-electric field is directly related to the attenuation of an acoustic wave. The quantity $\mathrm{Re}(\varrho_e v_d{}^*)$ is proportional to $\frac{1}{2}\mathrm{Re}(JE^*)$, the ohmic heat developed ($\varrho_e \propto J$ from (80) and $v_d \propto E$) which must be supplied by the acoustic wave which is thus attenuated. If this ohmic heat is the only cause of attenuation, then

$$\alpha P = \tfrac{1}{2}\mathrm{Re}(JE^*) = -\tfrac{1}{2}(v_s/\mu_e)\mathrm{Re}(\varrho_e v_d{}^*), \quad (86)$$

where α is the acoustic attenuation coefficient and P the energy flow density of the sound wave. (By definition, $\alpha = -(1/P)\eth P/\eth z$; since P is proportional to the square of the wave variables, $\alpha = -2\Delta k_i$.) In (86) we made use of $J = v_s\varrho_e$ (eq. 80) and $v_d = -\mu_e E$ (eq. I.2 in Table I).

From (85) and (86) the very simple relation between α and E_{ae} first given by G. Weinreich [33] follows directly:

$$\alpha = -nqv_s E_{ae}/P. \quad (87)$$

The very general reasoning used by Weinreich was as follows. A travelling acoustic wave represents not only a flow of energy but also a flow of momentum. Attenuation of the wave therefore implies a diminishing momentum flow: momentum is transferred to the 'cause of the attenuation' and hence a force acts on this 'cause', i.e. the charge carriers. The field equivalent to this force is the acousto-electric field. This reasoning shows that we are concerned here with a kind of *radiation pressure* exerted by the acoustic wave on the charge carriers.

Let us now make one or two further comments about Weinreich's relation.

Assuming that charge carriers of only one kind, of mobility μ_e, are responsible for the attenuation, the relation is very generally valid, even when the charge carriers have a non-zero constant drift velocity v_{d0}. This means that, as v_{d0} increases, E_{ae} changes its sign exactly at the instant when attenuation changes to amplification.

If the charge carriers do have a non-zero drift velocity, several field and current components come into play, as discussed on p. 337. It is not perhaps immediately clear whether only the part $\tfrac{1}{2}\mathrm{Re}(JE^*)$ of the ohmic heat is due to the acoustic wave (J and E are the wave variables) as is assumed in (86) (and not, for example, $J_{ae}E_0$). If this is not the case, then the derivation of (87) would no longer be valid. However, (86) can also be proved explicitly starting from the expression (78) for the attenuation and from the fact (not proved here) that the acoustic-energy flow density P can be represented by $\tfrac{1}{2}v_s cSS^*$. Since $\mathrm{Re}(JE^*)$ is equal to $\sigma_r EE^*$, (86) gives

$$\alpha = \frac{\sigma_r}{v_s c}\frac{EE^*}{SS^*}.$$

Substituting here the value for E/S from (77b), we arrive, using $\Delta k_i = -\tfrac{1}{2}\alpha$, directly at (78). This is an indirect proof of (86).

Much experimental work has been done on the acousto-electric effect, acoustic attenuation and the relationship between them, especially in CdS. In spite of what has been said above, Weinreich's relation is often not satisfied. For one thing, as v_{d0} increases, E_{ae} and α often do not go through zero at the same point. The reason for this is that some of the charge carriers are trapped and thus are not mobile although they still contribute to the space charge. These charge carriers cause absorption but they do not contribute to the acousto-electric current. From this reasoning [34] a generalization of Weinreich's relation can be derived by taking the *mobile* charge for ϱ_e in (85) and the *total* charge for ϱ_e in (86). Further analysis shows that measurement of the frequency dependence of E_{ae} gives information about the trapping mechanism.

Waves in two media with a common boundary

In bounded media, the bulk waves studied above can, in general, no longer occur by themselves because alone they do not satisfy the conditions imposed by the boundaries. At the free surface of a solid, for example, the shear stress tangential to the surface and the tensile stress normal to it are of course zero, whereas almost every simple sound wave involves such stresses. *Superposed* simple sound waves are however possible if when taken together they cause the surface stress to be zero. This indicates the way to tackle the present problem — the problem of two adjacent media: we should try to combine the bulk waves in such a way that the *boundary conditions* at the interface are always satisfied at all points of the interface. In this way we can describe phenomena such as refraction and reflection of waves and — our special interest here — surface waves. We shall first illustrate this approach by some simple examples.

Transmission, reflection, refraction

The junction of two transmission lines

In the transmission line of fig. 2 waves can be propagated with a voltage-current ratio V/I given by $\pm\sqrt{L/C}$. Or, more precisely, in the waves travelling to the right ($v > 0$), V/I is equal to $+\sqrt{L/C}$, the characteristic impedance Z of the line; in the waves travelling to the left ($v < 0$), V/I is negative, $V/I = -Z$. This follows directly from (3), (4) and (5). Let us now consider the reflection at the junction between two transmission lines of different characteristic impedances Z_1 and Z_2 (*fig. 28*), and consider what com-
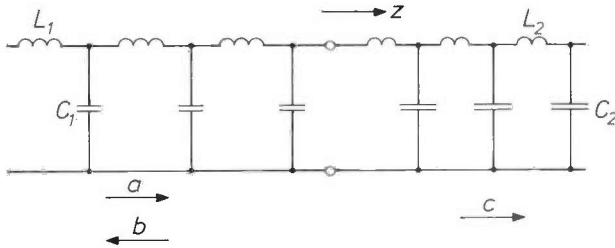
Fig. 28. Reflection and transmission at the junction of two transmission lines of characteristic impedances $Z_1 = \sqrt{L_1/C_1}$ and $Z_2 = \sqrt{L_2/C_2}$; a incident wave, b reflected wave, c transmitted wave.
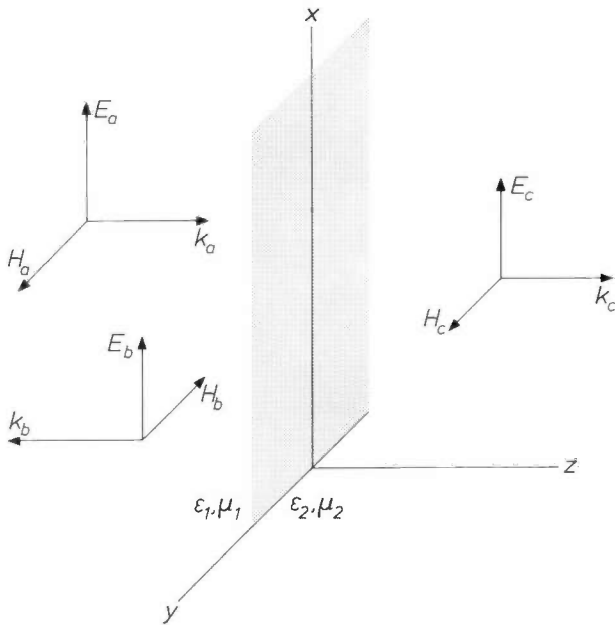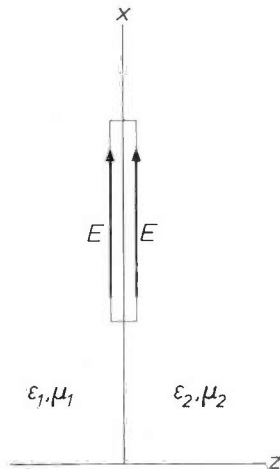

Fig. 29. Reflection and transmission of normally incident linearly polarized light at the interface between two media with intrinsic impedances $Z_1 = \sqrt{\mu_1/\varepsilon_1}$ and $Z_2 = \sqrt{\mu_2/\varepsilon_2}$; a incident wave, b reflected wave, c transmitted wave.

Fig. 30. Integration contour for E along the surface, for the proof that E is continuous at an interface (see text).

binations of incident (a), reflected (b), and transmitted (c) waves are possible. In this one-dimensional example the interface has degenerated to a junction. The boundary conditions require continuity of the current and voltage at the junction. If we take the junction to be at $z = 0$, then we must have (see eq. 2):

$$V_a \exp j\omega_a t + V_b \exp j\omega_b t = V_c \exp j\omega_c t,$$
$$I_a \exp j\omega_a t + I_b \exp j\omega_b t = I_c \exp j\omega_c t.$$

Since these equations must *always* be satisfied it follows, first that $\omega_a = \omega_b = \omega_c$ and secondly that

$$V_a + V_b = V_c,$$
$$V_a/Z_1 - V_b/Z_1 = V_c/Z_2. \tag{88}$$

From (88), the voltage reflection coefficient K_R (the ratio V_b/V_a) and the voltage transmission coefficient K_T (the ratio V_c/V_a) are, respectively:

$$K_R = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \qquad K_T = \frac{2Z_2}{Z_1 + Z_2}. \tag{89}$$

When the two lines are *matched*, i.e. $Z_1 = Z_2$, there is no reflection and complete transmission ($K_R = 0$, $K_T = 1$). When the first transmission line is open-circuited ($Z_2 = \infty$) or short-circuited ($Z_2 = 0$) there is complete reflection ($K_R = \pm 1$).

Normally incident light

When light falls normally on the interface between two isotropic transparent media (*fig. 29*) we again have the three waves incident (a), reflected (b) and transmitted (c). Suppose the light is linearly polarized. The boundary conditions in this case are that E and H are continuous at the interface. This can be shown for E by integrating E along an extended contour as in *fig. 30*. The result is curl E integrated over the enclosed area. If the loop of the contour is made infinitely thin, this surface integral is zero (curl E does not become infinite, see eq. 16) and therefore the contour integral is also zero, so that E must be the same on either side of the interface. The continuity of H is shown in a similar way.

The ratio E_x/H_y in each of the waves has an absolute value equal to the value of $\sqrt{\mu/\varepsilon}$ in the corresponding medium (see p. 317); the sign is positive for waves travelling to the right ($k > 0$), negative for waves

[33] G. Weinreich, Phys. Rev. 107, 317, 1957.
[34] C. A. A. J. Greebe, Physics Letters 4, 45, 1963.
    I. Uchida, T. Ishiguro, Y. Sasaki and T. Suzuki, J. Phys. Soc. Jap. 19, 674, 1964.
    P. D. Southgate and H. N. Spector, J. appl. Phys. 36, 3728, 1965.
    C. A. A. J. Greebe, IEEE Trans. SU-13, 54, 1966 and Philips Res. Repts. 21, 1, 1966.
    Y. Kikuchi, N. Chubachi and K. Iinuma, Jap. J. appl. Phys. 6, 1251, 1967.
    C. A. A. J. Greebe, 1968 Sendai Symp. on Acoustoelectronics, Sendai, Japan, p. 67.

travelling to the left $(k < 0)$; see (17a). Therefore

$$E_a + E_b = E_c,$$

$$E_a/\sqrt{\mu_1/\varepsilon_1} - E_b/\sqrt{\mu_1/\varepsilon_1} = E_c/\sqrt{\mu_2/\varepsilon_2}.$$

These equations have exactly the same form as (88). With the definition of the intrinsic impedance of a medium given on p. 317, $Z = \sqrt{\mu/\varepsilon}$, the reflection and transmission are given again by (89).

Normally incident sound

For sound in two adjoining elastic isotropic media (*fig. 31*) the situation is analogous to the foregoing. It is clear that, at the interface, T_{xz}, T_{yz}, T_{zz} and u_x, u_y, u_z must be continuous (interface perpendicular to the z-axis). Let us assume that in each medium only one component of T and u is involved. For waves travelling to the right we have (with $\omega/k = \sqrt{c/\varrho_m}$):

$$T = cS = c\partial_z u = -jkcu = -j\omega u \sqrt{c\varrho_m},$$

and for waves travelling to the left:

$$T = +j\omega u \sqrt{c\varrho_m}.$$

Therefore

$$T_a + T_b = T_c,$$

$$T_a/\sqrt{c_1\varrho_{m1}} - T_b/\sqrt{c_1\varrho_{m1}} = T_c/\sqrt{c_2\varrho_{m2}}.$$

The quantity $\sqrt{c\varrho_m}$ is called the *mechanical impedance*. The reflection and the transmission are again given by (89). The reasoning is valid for both transverse and longitudinal waves. Which component of T and of u and which constant c are relevant depends on the type of wave considered.

The close analogy between the above three cases might suggest that for normal incidence (89) is of very general validity. This is not the case. To obtain a result such as (89) with *one* transmitted and *one* reflected wave, these waves and their polarizations have to be matched to the incident wave. An example of a simple situation where this is no longer possible is the case of a linearly polarized beam of light that is normally incident on the interface between free space and an optically anisotropic uniaxial crystal when the optical axis lies *in* the interface but is not parallel or perpendicular to the plane of polarization of the light beam. *Two* beams of different velocities and different ratios E/H then arise in the crystal (see p. 318) so that (89) is no longer applicable. Of course, in this simple case, that incident beam can be resolved into two components with polarizations parallel and perpendicular to the optical axis, and (89) can then be applied to each of the two problems thus obtained.

Obliquely incident waves

In the three foregoing cases, the wave variables depended on z and t and thus, in the interface, only on t. Because the boundary conditions have to be satisfied *at all times* the waves must all have the same frequency. This virtually self-evident requirement was mentioned

Fig. 31. Reflection and transmission of normally incident sound at the interface between two elastic media *1* and *2*. It is assumed that for these waves each medium is sufficiently characterized by a density ϱ_m and a single elastic modulus c.
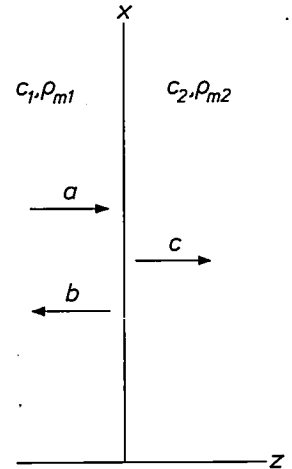
explicitly only in the first example. With obliquely incident waves, the wave variables also depend on the coordinates in the interface. Since the boundary conditions must be satisfied *at all times everywhere in the interface*, all waves combining at an interface must have the same periodicity both in *time* and in *place along the interface*. In the following we shall take the y-axis to be normal to the interface. Then all the waves of a combination must have the same ω, k_x and k_z. Once these three quantities are given, e.g. by the incident wave, then k_y is determined for each other wave of the combination by its dispersion relation (which is a relation between ω, k_x, k_y and k_z for that wave), and thus its direction of propagation is also determined. This is illustrated by the following example.

The laws of optical reflection and refraction

At the interface of two optically isotropic media, incident light is reflected and refracted (*fig. 32*). The velocity of light in medium *1* will be denoted by v_1, that in medium *2* by v_2. As stated above, the y-axis is taken to be perpendicular to the interface; the x,y-plane is taken as the plane of incidence (the plane
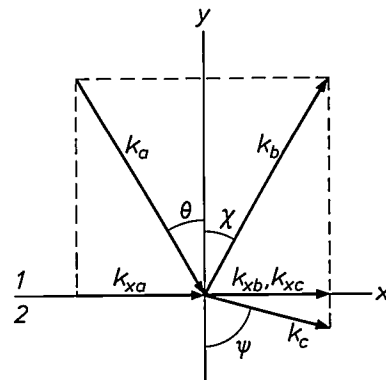

Fig. 32. Refraction and reflection of a light beam falling obliquely from medium *1* on to the interface $y = 0$ with a second medium 2; k_a, k_b and k_c wave vectors of the incident, reflected and refracted beams respectively.

containing k_a and the y-axis). Then $k_{za} = 0$. As a result k_{zb} and k_{zc} are also zero: the reflected beam (b) and the refracted beam (c) lie in the plane of incidence. The law of reflection, $\chi = \theta$, follows from fig. 32 by observing that not only are k_{xb} and k_{xa} equal, but k_b and k_a also, because the waves a and b are in the same isotropic medium and have the same frequency. Finally, Snell's law of refraction follows from the dispersion relations $k_a{}^2 = \omega^2/v_1{}^2$ and $k_c{}^2 = \omega^2/v_2{}^2$ for the waves a and c and from the fact that $k_{xc} = k_{xa}$:

$$\frac{\sin \psi}{\sin \theta} = \frac{k_{xc}/k_c}{k_{xa}/k_a} = \frac{v_2}{v_1}.$$

These laws are thus a direct consequence of the requirement for equal frequencies and equal wave-vector components along the interface. They do not of course represent all the information contained in the boundary conditions: as in the previous three cases, it is also possible to calculate *how much* light is reflected and how much refracted. This leads to Fresnel's laws, but we shall not consider these here.

In fig. 32 we assumed that the boundary conditions can be satisfied by *one* incident wave, *one* reflected wave and *one* refracted wave. For light waves in optically isotropic media, further investigation shows that this is indeed the case, but it is by no means a general rule. For longitudinal sound waves, for example, incident at a certain oblique angle, one reflected and one refracted longitudinal wave are not sufficient — it is also necessary to introduce *transverse* waves with different velocities and directions: *mode conversion* takes place at the interface. However, the wave vector of each of the waves is always completely determined by its dispersion relation and by the 'interface component' of the wave vector of the incident wave.

Total reflection; surface waves

Whatever the value of the angle of incidence θ (fig. 32), the number of variables and the number of boundary conditions remains the same, so that the *number* of waves necessary to satisfy the boundary conditions remains unchanged. In particular, when the angle of incidence is increased so far that total reflection occurs, *three* waves are still present. What then happens to the refracted wave can be seen by calculating k_{yc} from

$$k_{xc}{}^2 + k_{yc}{}^2 = k_c{}^2,$$

in which k_c is given by the dispersion relation for waves in medium 2:

$$k_c{}^2 = \omega^2/v_2{}^2,$$

and k_{xc} is given by k_{xa}:

$$k_{xc}{}^2 = k_{xa}{}^2 = (\omega^2/v_1{}^2) \sin^2 \theta.$$

If $\sin \theta$ becomes larger than v_1/v_2 — only possible for $v_2 > v_1$ — then $k_{xc}{}^2$ becomes larger than $k_c{}^2$. Hence k_{yc} is *imaginary*. The wave that was previously refracted now propagates *along* the surface (k_{xc} is *real*) but its amplitude in medium 2 decays exponentially at right angles to the surface. This means that the refracted wave has become a *surface wave*. This is why the reflection is total: no energy is carried away from the interface in medium 2. Total reflection thus implies a surface wave in medium 2.

It is clear that an imaginary k_{yc} (positive imaginary in fig. 32) implies that wave c transports no energy away from the boundary: no power is transmitted downwards *through* medium 2 because at large distances from the interface the wave has zero amplitude, and no energy is dissipated *in* the medium because our assumption was that it is lossless. We can calculate the situation explicitly as follows. The mean energy flow in the y-direction (see note [10]) is

$$\bar{S}_y = \tfrac{1}{2} \mathrm{Re}(E_z H_x{}^* - E_x H_z{}^*).$$

From Maxwell's equations

$$\mathrm{curl}\, H = \dot{D}, \quad \mathrm{curl}\, E = -\dot{B},$$

with

$$\partial_z = 0, \quad \partial_y = -jk_y, \quad \partial_t = j\omega, \quad B = \mu H, \quad D = \varepsilon E,$$

it follows that

$$H_x = k_y E_z/\omega\mu \quad \text{and} \quad E_x = -k_y H_z/\omega\varepsilon.$$

Substituting these values in the expression for $\bar{S}_y$ gives:

$$\bar{S}_y = \tfrac{1}{2}\mathrm{Re}\,[k_y{}^* E_z E_z{}^*/\omega\mu + k_y H_z H_z{}^*/\omega\varepsilon] =$$
$$= (k_{yr}/\omega\varepsilon\mu)\{\tfrac{1}{2}\varepsilon E_z E_z{}^* + \tfrac{1}{2}\mu H_z H_z{}^*\}.$$

Since the expression inside the curly brackets is real, k_y would have to have a non-zero real part if there is to be a mean energy flow in the y-direction.

In this demonstration it is assumed that the material is isotropic ($D = \varepsilon E$), that there are no losses (ε and μ real) and that ω is real. Under these conditions, the same is true for every other direction. For anisotropic media, on the other hand, it cannot be concluded that S has no component in a given direction if k has no real component in that direction. In fig. 6b, for example, S does have a component in the x-direction even though k_x is zero.

With an 'ordinary' wave, with real wave-vector components, the phase velocity ω/k_x in any direction (x) *other* than the direction of propagation is *greater* than the velocity of propagation ω/k, since $k_x \leqslant k$. For a surface wave, on the other hand, as follows from the foregoing, $k_x > k$: the surface wave is propagated *more slowly* than the corresponding bulk wave. Surface waves are therefore said to be 'slow'.
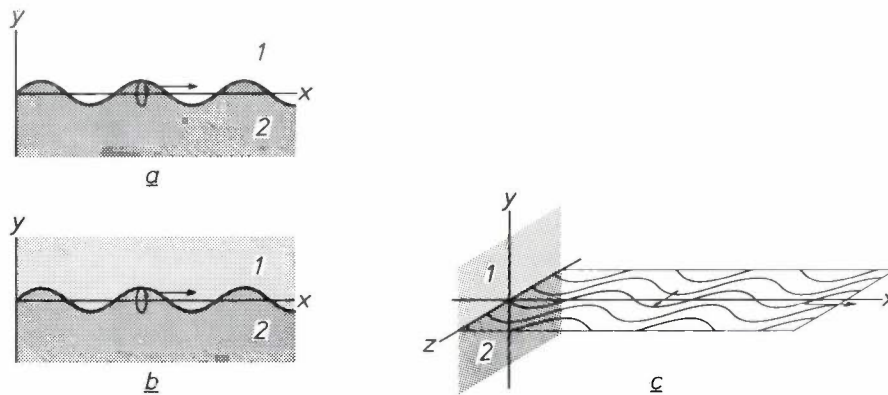
The phenomenon of total reflection can thus be summarized as follows. As the angle of incidence θ is increased, the phase velocity along the surface of all the waves involved decreases. When this velocity becomes smaller than the velocity of bulk waves in medium 2, the wave in medium 2 becomes a surface wave.

The surface wave is excited by the incident wave. In certain circumstances, however, *independent freely*

propagating surface waves are possible. This is the case when *all* the waves necessary to satisfy the boundary conditions are surface waves, i.e. waves that have a real wave-vector component along the surface (the same for all of them) and for which the wave variables decay exponentially with distance from the surface. Such waves form the subject of the next section.

Surface waves

As was mentioned in the introduction, acoustic surface waves on the free surface of a piezoelectric material offer interesting possibilities in certain fields of electronics. Such waves carry an electric field that extends beyond the boundary of the medium so that they can be generated, detected, amplified or otherwise processed electrically anywhere on the surface.

type of surface wave is possible, the Love wave (fig. 33c). The particle movement is confined here to the z-direction. In this case, however, if medium *1* is made less dense, the penetration depth of the waves in medium *2* becomes greater, and in the limit the Love wave does not remain a surface wave but degenerates to a *bulk* wave, propagating parallel to the surface without being perturbed by it.

In piezoelectric media the situation differs from the foregoing only in this last respect. Rayleigh-like, Stoneley-like and Love-like waves are all possible, now carrying electric polarizations and fields. The Rayleigh waves can again be regarded as a special case of the Stoneley waves. But the Love wave no longer reduces to a bulk wave when the density of one medium becomes zero as it does in the case of a non-piezoelectric medium. Instead, it remains a surface wave — at least

Fig. 33. Elastic interface waves. *a*) The Rayleigh wave propagates along the free surface of an isotropic elastic medium *2* (*1* is free space); in the coordinate system used (*y*-axis normal to surface, *x*-axis in propagation direction of wave) the particle movement is in the *x,y*-plane. Waves are also possible along the interface between two bonded isotropic elastic media (*1* and *2*), e.g. Stoneley waves (*b*) with particle movement in the *x,y*-plane and Love waves (*c*) with particle movement in the *z*-direction. If the density of medium *1* tends to zero, we find the Rayleigh wave as a special case of the Stoneley wave; the Love wave, however, becomes a bulk wave (the penetration depth in the medium *2* becomes infinite). If, however, medium *2* is piezoelectric (and some other conditions are satisfied, see text), the Love wave remains a surface wave if the density of medium *1* becomes infinitely small; this limiting case is the Bleustein-Gulyaev wave.

Let us for a moment consider isotropic solid media that are *not* piezoelectric. The well known *Rayleigh wave* can be propagated on the free surface of such a medium (*fig. 33a*). In a coordinate system in which the surface is perpendicular to the *y*-axis, and the surface wave propagates in the *x*-direction, the particles move in ellipses in the *x,y*-plane for a Rayleigh wave. The Rayleigh wave is a special case of the *Stoneley wave* (fig. 33b). This can occur at the common boundary of *two* elastically different solid media bonded to each other — at least, when the elastic moduli and the densities satisfy certain conditions. Here again the particle movement is in the *x,y*-plane. If the density of medium *1* tends to zero, the Stoneley wave reduces to the Rayleigh wave.

At the interface of two bonded media yet another

in the case of a piezoelectric medium with an axis of rotation or a six-fold axis of symmetry lying in the surface and direction of propagation along the surface perpendicular to this axis. This is the *Bleustein-Gulyaev wave* [35] discovered theoretically in 1969. This wave is a surface wave *only* because of the piezoelectric nature of the medium: if the piezoelectric coupling were zero, the penetration depth would become infinite. The existence of the Bleustein-Gulyaev wave has been confirmed by many experiments [36], made, however, in configurations of a complexity far beyond that suggested by the simple situation of fig. 33c.

Rayleigh, Stoneley and Love waves are rather complicated and are not easily treated mathematically, especially in piezoelectric materials. The Bleustein-Gulyaev wave, on the other hand, with its particle

movement confined to the z-direction, is a relatively simple example of an acoustic surface wave on a piezoelectric medium. We shall now consider this wave in somewhat more detail. The treatment is different from that usually given but is perhaps more instructive [37].

Bleustein-Gulyaev waves

Let us consider a piezoelectric medium bounded by the plane $y = 0$, outside which there is free space (*fig. 34*; $y < 0$ medium, $y > 0$ free space). We shall show that under certain conditions combinations of waves exist in the piezoelectric medium and in the free space that consist entirely of surface waves and in combination satisfy the boundary conditions. These combinations are called Bleustein-Gulyaev waves. The common ω and k_x of the waves are assumed real and positive, so that the waves in fig. 34 are propagated to the right.

The calculation is given in condensed form in *Table II* (opposite p. 348). The approach and the result of these calculations are roughly as follows. We put certain restrictions on the piezoelectric medium and on the waves to be considered. An introductory calculation first shows that the boundary conditions can be satisfied under these restrictions by the superposition of *four* waves — not all surface waves as yet — of given real positive ω and k_x (see *fig. 35*, upper diagram). These four waves are: *a*) a surface wave in free space; *b*) a surface wave in the piezoelectric medium; *c*) an incident wave and *d*) a reflected wave of 'stiffened sound' (see p. 336), at least, if the given k_x is smaller than the wave vector k_s of 'stiffened sound' in the piezoelectric material. The waves *c* and *d* have equal but opposite real k_y. Denoting the amplitudes of the potentials of these four waves by ϕ_a, ϕ_b, ϕ_c and ϕ_d, respectively, then ϕ_a, ϕ_b and ϕ_d can be expressed in terms of ϕ_c because there are *three* independent boundary conditions. The problem of the reflection of 'stiffened sound' at a free surface is thus solved. However, this was not our problem: we were in search of a surface wave.

As for the refracted wave in the refraction of light (p. 343), we may now enquire what happens to the waves *c* and *d* if for any reason k_x should become larger than k_s. The wave-vector components k_y of *c* and *d* then become imaginary. They remain however of opposite sign and thus we do find a 'well behaved' surface wave propagating in the x-direction whose amplitude falls off exponentially with distance from the surface (let this be *c*); however, we also find a wave (*d*) that would again propagate in the x-direction, but whose amplitude would *increase* exponentially with distance from the surface, and which is therefore unacceptable.

The combination of waves *a*, *b*, *c* and *d* necessary to satisfy the boundary conditions therefore no longer forms an acceptable solution.

Nevertheless a freely propagating surface wave can be found, since for any real positive ω we can find a real, positive k_x that is larger than k_s and makes ϕ_d equal to zero. The unacceptable wave thus vanishes from the scene (fig. 35, lower diagram). In other
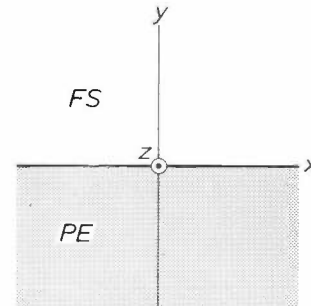

Fig. 34. Coordinate system used for the derivation of the Bleustein-Gulyaev wave; $y = 0$ is the interface between the piezoelectric medium *PE* ($y < 0$) and free space *FS* ($y > 0$). The wave propagation is in the x,y-plane, the particle movement in the z-direction. In this coordinate system the piezoelectric material must have an array of coefficients of the form M in Table II (opposite p. 349).
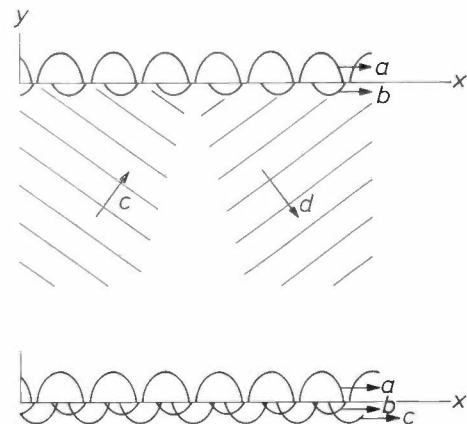

Fig. 35. *Above*: a wave of stiffened sound (*c*) incident from the piezoelectric material on the interface gives rise to a reflected wave (*d*) and two surface waves, one in the free space (*a*) and the other in the piezoelectric medium (*b*). The common k_x of the four waves is smaller than the wave vector k_s of stiffened sound in the given piezoelectric material. If k_x becomes larger than k_s then the wave-vector components k_y of (*c*) and (*d*) become imaginary; they remain opposite in sign. Both of these waves thus propagate along the surface, and while the amplitude of one of them decreases, that of the other increases exponentially away from the surface. At a certain value of k_x ($> k_s$), however, the unacceptable wave (*d*), whose amplitude increases away from the surface, vanishes from the solution. The solution is then the Bleustein-Gulyaev wave (*lower diagram*).

[35] J. L. Bleustein, Appl. Phys. Letters **13**, 412, 1968.
      Yu. V. Gulyaev, JETP Letters **9**, 37, 1969.
[36] See for example P. A. van Dalen and C. A. A. J. Greebe, Philips Res. Repts. **27**, 340, 1972, and P. A. van Dalen, Philips Res. Repts. **27**, 323, 1972.
[37] A more comprehensive survey of possible surface and interface waves and further literature references are given in the article by R. M. White [3].

words, when we attempt to satisfy the boundary conditions for a given ω with the *three* waves a, b and c, we find that this is possible with one value of k_x that is larger than k_s, so that c is also a surface wave. This combination of three surface waves is the Bleustein-Gulyaev wave, and the relation found between k_x and ω is its dispersion relation.

In our description of the Bleustein-Gulyaev wave we thus allow an incident wave and a reflected wave (real k_y) to change into two surface waves (imaginary k_y), one decreasing and the other increasing exponentially in amplitude with distance from the surface. We then cause the unacceptable increasing wave to vanish by choosing a suitable value $k_x(\omega)$ for k_x. We note that this approach is of general application to surface waves: the Bleustein-Gulyaev wave merely serves here as an example.

We shall now look into one or two details of the calculation. We impose the following restrictions (see *R1-R4* in Table II): the piezoelectric material must have a certain symmetry and a certain orientation (*R4*); the waves must be 'slow' (*R2*), must propagate in the x,y-plane (*R1*) and must involve particle displacement in the piezoelectric medium in the z-direction only (*R3*). The waves must be slow in the sense that only electrostatic effects and no electrodynamic effects arise (see also p. 336); this means that the electric field can be derived from a potential ϕ ($E = -\mathrm{grad}\ \phi$).

For the waves indicated in Table II as *FSL*, which can occur in free space under the restrictions *R1* and *R2*, $k^2 = 0$ for all ω. The wave variables thus satisfy the two-dimensional Laplace equation, for $k^2 = k_x^2 + k_y^2 = -(\partial_x^2 + \partial_y^2) = -\nabla^2$. These are in fact static field distributions which can propagate at an arbitrary (but not too high) velocity, provided that the boundary conditions at the surface are satisfied. For k_x real, k_y is evidently imaginary: $k_y = \pm\ \mathrm{j}k_x$. The wave to be used for combination (wave a in list L, Table II) must decrease exponentially with distance from the surface (i.e. upwards in fig. 34) so that for the given k_x, the k_y of the wave must be equal to $-\mathrm{j}k_x$. Lines of force and equipotential surfaces of this *Laplace wave* are given in *fig. 36*.

Restrictions on the symmetry and orientation (*R4*) of the piezoelectric medium are imposed in the array marked *M* in Table II, which gives the coefficients by means of which the variables $T_1, T_2, \ldots D_x, D_y, D_z$ are expressed in terms of the variables $S_1, S_2, \ldots E_x, E_y, E_z$. The many zeros imply a high symmetry and a simple orientation; the coefficients indicated by points are not relevant to the present discussion. Materials with a six-fold or rotational axis of symmetry are examples in which the array of coefficients can take the form *M*. The restrictions *R1*, *R2* and *R3* have already

limited the components of S and E to be taken into account here to S_4, S_5, E_x and E_y and now because of the form of the array the components of T and D are limited to T_4, T_5, D_x and D_y.

As a result of the foregoing, there only remain two differential equations for the waves in the piezoelectric material: a mechanical equation of motion and one of Maxwell's equations. The resultant dispersion relation represents two waves — again, Laplace waves (*PL*), with $k^2 = 0$ and further 'stiffened sound' (*PS*), with the dispersion relation (*DS*). This reduction of the possible waves to two simple types is of course a consequence of the simplifications introduced by (*M*). The Laplace wave to be used for combination (wave b in list L, Table II) must decrease with distance from the surface, i.e. downwards in fig. 34, and thus has $k_y = \mathrm{j}k_x$. For the incident sound wave (c in list L), $k_y > 0$; the reflected wave (d) has a negative k_y. In the list L of possible waves, only wave variables are included that enter into the boundary conditions.

The boundary conditions *B1*, *B2*, *B3* express, firstly, that the boundary $y = 0$ is a mechanically free surface. The shear stresses ($T_{xy} = T_6$ and $T_{yz} = T_4$) along the surface and the tensile stress ($T_{yy} = T_2$) normal to it are thus zero. Two of the boundary conditions ($T_2 = 0$, $T_6 = 0$) are automatically satisfied in all the waves considered, and therefore only one mechanical boundary condition remains (*B1*): $T_4 = 0$. The electrical boundary conditions (*B2*, *B3*) stating that ϕ and D_y must be continuous at the surface, follow because at the surface the field must be derivable from a potential and there is no source of D, i.e. no charge. We note that before the introduction of the boundary conditions, the factor $\exp\ \mathrm{j}(\omega t - k_x x - k_y y)$ in each wave plays no part — it cancels out in all equations. However, when various waves are combined, *different* factors $\exp\ (-\mathrm{j}k_y y)$ are involved. At the boundary these factors again vanish since $y = 0$ there.

Substitution of the wave variables from list L in the boundary conditions (*B*) yields, after eliminating ϕ_a and ϕ_b, the relation (*A*) between ϕ_c and ϕ_d. Therefore we arrive at the two conclusions mentioned earlier:
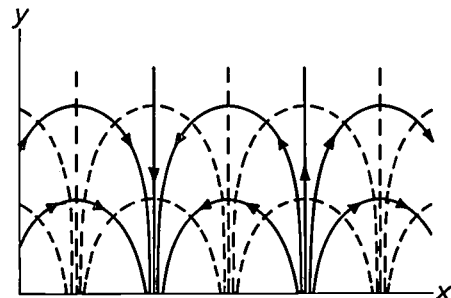

Fig. 36. Lines of force (solid curves) and cross-sections of equipotentials (dashed) of the Laplace wave which decays exponentially for $y \rightarrow +\infty$.

I. A wave of 'stiffened sound' incident at a certain angle imposes a k_x and thus determines indirectly all the k_y's; since k_x is smaller than k_s, k_{yc} and k_{yd} turn out to be real. Using (A) in Table II, all the wave variables can be expressed in terms of ϕ_c. The angle of reflection is equal to the angle of incidence $(k_{yd} = -k_{yc})$. The reflected wave has the same amplitude as the incident wave $(|\phi_d| = |\phi_c|$, because the coefficients of ϕ_c and ϕ_d in (A) are the complex conjugates of one another).

II. From (A) we can produce a pure surface wave, the Bleustein-Gulyaev wave, because wave d vanishes $(\phi_d = 0)$ for the value of k_{yc} given in (K) (Table II); this value of k_{yc} is positive imaginary and thus c becomes a 'well behaved' surface wave. Substituting this value of k_{yc} in the identity $k_x{}^2 + k_y{}^2 \equiv k_s{}^2$ yields expression (DBG) in Table II, the dispersion relation for the Bleustein-Gulyaev wave. The velocity ω/k_x is independent of ω: the Bleustein-Gulyaev wave is therefore dispersionless.

Notes on the Bleustein-Gulyaev wave

In the foregoing derivation of the Bleustein-Gulyaev wave it was not necessary — as it was in other problems discussed in this article — to assume that $e^2/\varepsilon c$ was much less than unity. This underlines once more the relative simplicity of the Bleustein-Gulyaev wave. We assumed only that the wave was 'slow' and the dispersion relation (DBG) shows that this is always the case except when $e^2/\varepsilon c \gg 1$, a situation that in fact never arises.

From (K) it can be seen that the penetration depth increases as $e^2/\varepsilon c$ decreases and becomes infinite, as mentioned earlier, in the limit $e^2/\varepsilon c \to 0$; the surface wave has then degenerated to a bulk wave. In contrast to this, the penetration depth of a Rayleigh-like wave always remains of the same order as the wavelength along the surface. Bearing in mind that, in practical applications, the great attraction of surface waves lies in their small penetration depth, it can be seen that the Bleustein-Gulyaev wave has its limitations; only when $e^2/\varepsilon c$ is approximately unity can it match the Rayleigh wave in this respect.

The factor $\varepsilon_0/(\varepsilon_0 + \varepsilon)$ in (K) also tends to make $|k_{yc}|$ small and hence the penetration depth large because ε is often much larger than ε_0 in piezoelectric materials. In the above the empty space behaves only as a medium of permittivity ε_0 with no mechanical effect on the surface. If the empty space is replaced by another medium with a large permittivity but still with no mechanical effect on the surface, then the penetration depth can be reduced. For this reason the surface is sometimes coated with a thin film of metal ($\varepsilon \approx \infty$). Further analysis shows that the wave a then vanishes

$(\phi_a = 0)$; there is virtually no penetration in the metal so that the metal film can be so thin that its mechanical effect is negligible. Strictly speaking, we are then dealing with a wave problem in a single bounded medium, the piezoelectric material, but with different electrical boundary conditions.

Amplification of surface waves; the effect of a transverse magnetic field

An acoustic surface wave on a piezoelectric material is accompanied by an electric field that extends beyond the surface and propagates there as a Laplace wave. The wave can be excited, detected, directed, amplified and its dispersion relation changed by means of this electric field. Here we shall only examine the amplification of such waves and we shall show that with a transverse magnetic field the amplification can be enhanced.

From the foregoing it is fairly evident how to set about amplifying a surface wave on a piezoelectric material: a semiconductor is placed against the piezoelectric material and a current is passed through it of such a value that the drift velocity of the electrons is higher than the wave velocity. We shall see presently that this should work; that it does work was first shown experimentally by J. H. Collins et al. [6]. Compared with the bulk-wave acoustic amplifier (see p. 337) the present configuration has the advantage that the piezoelectric material and the semiconductor can be selected independently, so that an optimum choice can be made. This is one of the reasons why the surface-wave amplifier seems to be a step nearer to practical applications than the bulk-wave amplifier. Nevertheless, there are still great difficulties associated with the surface-wave amplifier. In the first place, the required drift field is very high (this point is illustrated in fig. 27); this means that it is generally necessary to dissipate undesirable large amounts of power unless certain precautions are taken that present practical difficulties.

Secondly there is the problem of the electric coupling between the piezoelectric material and the semiconductor. In a Rayleigh wave, the surface particles move primarily in the y-direction (see fig. 33). This wave is therefore very sensitive to mechanical contact with the surface. To avoid this difficulty it can be arranged, for example, to have a gap between the surfaces of the piezoelectric material and the semiconductor. This gap, however, must be very small (a small fraction of the wavelength), for otherwise there will be no electric coupling. The Bleustein-Gulyaev wave, in which there is particle movement in the z-direction only, is better in this respect: it is very little affected by the presence of a substance such as a liquid (which attenuates a Rayleigh wave strongly), and a dielectric

liquid of high permittivity in the gap results in a strong electric coupling. This, however, may give rise to new difficulties, such as corrosion of the surfaces by the liquid.

We shall now indicate briefly how acoustic surface waves are amplified and how this effect can be enhanced by a transverse field [38]. In *fig. 37a*, *1* denotes the semiconductor and *2* the piezoelectric material. In the semiconductor the slow wave consists of a Laplace wave (see fig. 36); associated with the field E, there is the bulk current $J_b = \sigma_0 E$. In a Laplace wave there can be no charge fluctuations *in* the material since in the Laplace wave div $D = -\varepsilon \nabla^2 \phi = 0$. The current that flows towards and away from the surface gives, however, an alternating *surface* charge. In fig. 37a it is assumed that the d.c. drift velocity of the electrons is zero and that the Laplace wave, including the surface-charge pattern, travels to the right at velocity v_s. The maxima of positive and negative surface charge must assume, with respect to the field pattern, the phases as shown: on the right (front) of the positive charge maxima, the charge must increase, so the current must be directed towards these points. In fig. 37a the field and the current are in phase: energy is therefore dissipated.

If the electron gas as a whole is now made to move through the semiconductor at a drift velocity v_{d0}, the pattern of alternate surface charges itself represents — quite independently of the alternating field E — an *a.c. surface current J_s* (fig. 37b). A smaller bulk current will now maintain the same surface charge; or the same alternating field E and bulk current J_b will now give rise to a surface-charge wave of larger amplitude. When v_{d0} is equal to v_s no bulk a.c. current will be required to maintain a surface-charge wave. (Diffusion and trapping of charge carriers are neglected here.) If, finally, v_{d0} becomes larger than v_s, charge has to be *removed* from the front of each positive charge maximum by the bulk a.c. current (fig. 37c). Field and current now have opposite phase and hence energy is *supplied* to the wave.

It will be clear that the operation of the surface-wave amplifier is rather similar to that of the bulk-wave amplifier. In the surface-wave amplifier, however, advantage can be taken of a transverse magnetic field in a way that has no parallel in the case of bulk acoustic waves. We shall attempt to explain this by means of the equation

$$\frac{\Delta k_i}{k} = -\frac{1}{2}\frac{e^2}{\varepsilon c}\frac{\varepsilon}{\varepsilon + \sigma_i/\omega}\frac{(\varepsilon + \sigma_i/\omega)\,\sigma_r/\omega}{(\varepsilon + \sigma_i/\omega)^2 + (\sigma_r/\omega)^2}, \quad (90)$$

which is simply (78) in a slightly different form. This relation derived for phenomena in the bulk, is not of course exactly valid in this form for surface waves, but

a very similar relation is valid and we use (90) to indicate qualitatively the effect of a transverse magnetic field [39]. The last factor in (90) is of the form $pq/(p^2 + q^2)$ and it therefore has values between $+\frac{1}{2}$ and $-\frac{1}{2}$. The 'best' value, $-\frac{1}{2}$, can always be achieved, for arbitrary values of ε and σ_i, by giving σ_r, via the drift velocity v_{d0}, a suitable (negative) value. From the other factors in (90) it can be seen that the maximum amplification obtainable in this way could be increased if ϱ_i could be made negative, i.e. if the phase difference between J and E could be changed in a certain way. In the bulk-wave amplifier there is no way of doing this. In the surface-wave amplifier, however, it can be done by means of a transverse magnetic field. To make this clear, fig. 37d shows the extreme case of a magnetic field so large that there is a Hall angle of 90° between J and E. The figure shows that for a travelling Laplace
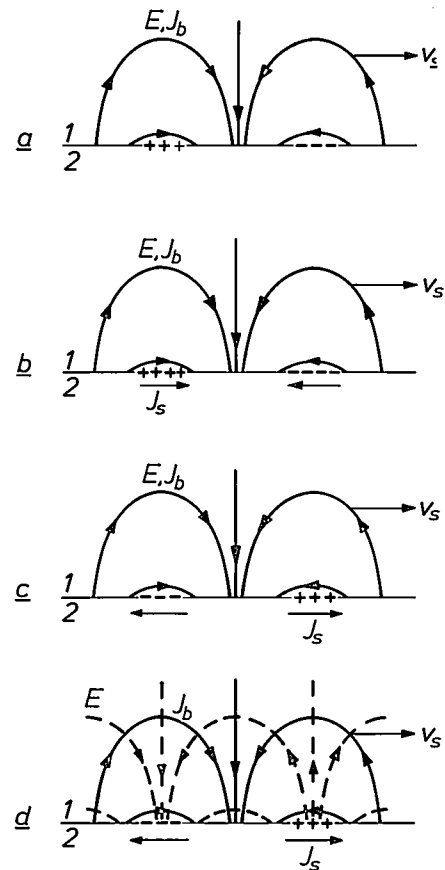

Fig. 37. Operation of the acoustic surface-wave amplifier and the effect of a transverse magnetic field. *1* semiconductor, *2* piezoelectric material. In the semiconductor, the field E of the Laplace wave (velocity v_s) carries with it a bulk a.c. current J_b and an alternating surface charge. For a given alternating field, the amplitude and phase of the surface-charge wave are dependent on the mean drift velocity v_{d0} of the electrons in the semiconductor (see text). *a*) $v_{d0} = 0$; *b*) $0 < v_{d0} < v_s$; *c*) $v_{d0} > v_s$. In case (c) amplification occurs. J_s is the surface a.c. current directly associated with the movement of the surface charges with the electron gas as a whole at velocity v_{d0}. *d*) When a large magnetic field is present E and J_b are at right angles to one another which means, for the travelling Laplace wave, a phase difference of 90° between J and E. Such a phase difference can lead to greater amplification.

wave this corresponds to a phase difference of 90° between J and E. It is also clear that smaller phase differences are introduced by smaller fields and that the sign of the phase difference is determined by the polarity of the magnetic field.

The effect of a transverse field on the amplification has been experimentally confirmed for Rayleigh waves by J. Wolter [40]. Some particulars concerning this experiment are given in *fig. 38*.

Fig. 38. Effect of a transverse magnetic field on the amplification and attenuation of Rayleigh waves, *a*) experimental *b*) theoretical, after J. Wolter [40]. The amplification in dB is plotted in (*a*) as a function of the applied drift field E_0, for various values of the transverse magnetic flux density B; in (*b*) the amplification is plotted as a function of the drift velocity v_{d0} for various values of μB (μ = mobility of the electrons in the semiconductor). The Rayleigh waves have a frequency of 50 MHz and are propagated on the surface of a plate of the piezoelectric material LiNbO$_3$. They are amplified or attenuated by the electrons moving in a silicon plate (type N, 150 Ωcm, 2 mm long×200 μm thick) on that surface. The values of μB and v_{d0} in (*b*) are matched (i.e. chosen so that (*b*) fits (*a*) as well as possible). The silicon was pressed against the LiNbO$_3$ so that the two media were in intimate mechanical contact at one or two places (the Rayleigh wave is therefore somewhat attenuated). It was established optically that the gap was less than 0.1 μm over a length of about 0.4 mm. Only over this length was the silicon effective. The calculated amplification and attenuation were in reasonable agreement with experiment.

Summary. A survey is given of freely propagating electromagnetic, elastic and electro-elastic waves, the accent falling on certain types of wave that have attracted attention in electronics and solid-state physics during the last decade or so. These include helicon waves, amplifying acoustic waves and electroacoustic surface waves; several more conventional waves are also discussed by way of introduction.

The dispersion relation and structure of a travelling wave in an unbounded homogeneous medium follow from the differential equations (assumed to be linear and homogeneous) for the appropriate variables. By allowing the frequency ω or the wave vector k to be complex, the attenuation or amplification of the wave in time or space can be represented.

Starting with Maxwell's equations, electromagnetic waves (light) are discussed, first in free space and then in other non-conducting media, whose particular properties can be formally expressed in terms of a permittivity which may be complex and may be a tensor. In this way double refraction and rotation of the plane of polarization, as in Faraday rotation, are described. Application of the results to conductors (with a permittivity that formally includes the conductivity) leads to wave phenomena in metals such as the skin effect and, for very high conductivities and very high magnetic fields, helicon waves. The latter are virtually unattenuated circularly polarized waves (simplest case) with a very strong dispersion so that, for low frequencies, their velocity is extremely low (e.g. 10 cm/s at about 20 Hz).

At the plasma frequency of a conductor, transverse EM waves change from cut-off waves (k imaginary) into travelling waves and longitudinal waves of arbitrary wavelength are possible. Examples of other longitudinal EM wave phenomena are dielectric relaxation (ω imaginary) and Debye-Hückel screening (k imaginary, ω zero).

Of the purely elastic waves, in which only mechanical variables are involved (displacements, deformations and stresses) only those in isotropic media are discussed.

In media where one set of variables is weakly coupled with another set, the possible waves can often be considered as if they were two separate coupled waves. The coupling between elastic and EM variables in piezoelectric materials leads to 'near-light' and 'near-sound' (electric transverse waves) and to 'stiffened sound' (electric longitudinal waves). In piezoelectric semiconductors the stiffening is complex. This leads, for real ω, to an 'amplifying' wave (amplification) if the real part of the complex conductivity becomes negative. This can be achieved by means of a constant electric field that gives the electrons a certain drift velocity in the semiconductor (acoustic amplifier). The acoustic attenuation or amplification coefficient is directly related to the acousto-electric field (Weinreich's relation).

In two media with a common boundary the possible wave configurations are found by superposition of simple waves with due regard to the boundary conditions. In this way the reflection and transmission coefficients can be found for waves at the junction of two transmission lines or for light or sound incident on an interface. All component waves must have the same frequency and the same wave-vector component along the common boundary. The laws of reflection and refraction are a direct consequence of these requirements. When total reflection occurs one of the component waves becomes a surface wave.

In a freely propagating surface wave, *each* component wave is a surface wave. As an example, the Bleustein-Gulyaev wave, known since 1968, is described. This is a surface wave by virtue of the piezoelectric property of the medium. This wave is closely related to the reflection of 'stiffened sound' at the surface of a piezoelectric medium. Acoustic surface waves on a piezoelectric medium can be amplified by a drift current in an adjacent semiconductor. A transverse magnetic field can enhance this amplification in an interesting way.

[38] A fuller discussion of these and related subjects and further references are given in: C. A. A. J. Greebe, P. A. van Dalen, T. J. B. Swanenburg and J. Wolter, Electric coupling properties of acoustic and electric surface waves, Physics Reports 1C, 235-268, 1971.

[39] In many problems the description can be made in terms of either a complex σ or a complex ε (see for example eq. 33). Here we use a complex σ. In the article of note [38] a complex ε was used.

[40] J. Wolter, Physics Letters 34A, 87, 1971.

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, 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 Forschungslaboratorium Aachen GmbH, Weißhausstraße, 51 Aachen, Germany *A*

Philips Forschungslaboratorium Hamburg GmbH, 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, Netherlands.

V. Belevitch: On the realizability of non-rational positive real functions.
Int. J. Circuit Theory & Appl. **1**, 17-30, 1973 (No. 1). *B*

H. J. van den Berg & A. J. Luitingh: Reproducibility and irreproducibility of etching time in freeze-etch experiments.
Cytobiologie **7**, 101-104, 1973 (No. 1). *E*

J. Bloem (Philips Semiconductor Development Laboratory, Nijmegen): High chemical vapour deposition rates of epitaxial silicon layers.
J. Crystal Growth **18**, 70-76, 1973 (No. 1). *E*

P. W. J. M. Boumans: Spektralanalysen. Optische Atomspektroskopie.
Techn. Rdsch. **63**, No. 37, 49-53 & No. 43, 33-37, 1971. *E*

K. H. J. Buschow, A. M. van Diepen & H. W. de Wijn (State University of Utrecht): Evidence for RKKY-type interaction in intermetallics, as derived from magnetic dilution of GdPdIn with Y or Th.
Solid State Comm. **12**, 417-420, 1973 (No. 5). *E*

M. C. W. van Buul & 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 (No. 4). *E*

K. L. Bye, P. W. Whipps & E. T. Keve: High internal bias fields in TGS (L-alanine).
Ferroelectrics **4**, 253-256, 1972/73 (No. 4). *M*

F. M. A. Carpay: Theory of and experiments on aligned lamellar eutectoid transformation.
Reactivity of solids, Proc. 7th Int. Symp., Bristol 1972, pp. 612-622. *E*

F. M. A. Carpay: Aligned composite materials obtained by solid state decomposition.
J. Crystal Growth **18**, 124-128, 1973 (No. 2). *E*

V. Chalmeton: Neutron radiography.
Acta Electronica **16**, 73-84, 1973 (No. 1). *(Also in French.)* *L*

J. W. Chamberlayne & B. Gibson: Magnetic materials for integrated cores.
IEEE Trans. **MAG-8**, 759-764, 1972 (No. 4). *M*

T. A. C. M. Claasen, W. F. G. Mecklenbräuker & J. B. H. Peek: Some remarks on the classification of limit cycles in digital filters.
Philips Res. Repts. **28**, 297-305, 1973 (No. 4). *E*

J. A. Clarke & E. C. Yeadon (Mullard Ltd, Mitcham, Surrey): Measurement of the modulation transfer function of channel image intensifiers.
Acta Electronica **16**, 33-41, 1973 (No. 1). *(Also in French.)* *M*

G. Clément & C. Loty: The use of channel plate electron multipliers in cathode-ray tubes.
Acta Electronica **16**, 101-111, 1973 (No. 1). *(Also in French.)* *L*

J. Cornet & D. Rossier: Phase diagram and out-of-equilibrium properties of melts in the As-Te system.
Mat. Res. Bull. **8**, 9-20, 1973 (No. 1). *L*

C. Z. van Doorn: On the magnetic threshold for the alignment of a twisted nematic crystal.
Physics Letters **42A**, 537-539, 1973 (No. 7). *E*

H. Dormont: Modèle théorique de système à avalanche conduisant à une étude du bruit.
Le bruit de fond des composants actifs semi-conducteurs, Coll. Int. C.N.R.S. No. 204, Toulouse 1971, pp. 189-192; 1972. *L*

D. L. Emberson (Mullard Ltd, Mitcham, Surrey) & **R. T. Holmshaw:** The design and performance of an inverting channel image intensifier.
Acta Electronica **16**, 23-32, 1973 (No. 1). *(Also in French.)* *M*

Table II. Derivation of the Bleustein-Gulyaev wave

R1. Restriction to waves propagating in the x,y-plane: $\partial_z = 0$

R2. Restriction to slow waves: $E = -\operatorname{grad}\phi \rightarrow$ $\begin{cases} E_x = jk_x\phi \\ E_y = jk_y\phi \\ E_z = 0 \end{cases}$

Waves in free space	Waves in piezoelectric medium	

Waves in piezoelectric medium

R3. Restriction to displacement in the z-direction: $\begin{cases} u_x = 0 \\ u_y = 0 \end{cases} \rightarrow \begin{cases} S_4 = -jk_y u_z \\ S_5 = -jk_x u_z \\ \text{other } S \text{ are} \\ \text{zero} \end{cases}$

R4. Piezoelectric coefficients:

	S_1	S_2	S_3	S_4	S_5	S_6	E_x	E_y	E_z		(M)
T_1	:	.	.	0	0	.	0	0	.
T_2	:	.	.	0	0	.	0	0	.
T_3	:	.	.	0	0	.	0	0	.
T_4	0	0	0	c	0	0	0	$-e$	0
T_5	0	0	0	0	c	0	$-e$	0	0
T_6	:	.	.	0	0	.	0	0	.
D_x	0	0	0	0	e	0	ε	0	0
D_y	0	0	0	e	0	0	0	ε	0
D_z	.	.	.	0	0	.	0	0	.

$D = \varepsilon_0 E$

$\begin{cases} T_4 = -jk_y(cu_z + e\phi) \\ T_5 = -jk_x(cu_z + e\phi) \\ D_x = -jk_x(eu_z - \varepsilon\phi) \\ D_y = -jk_y(eu_z - \varepsilon\phi) \\ \text{other } T \text{ and } D \text{ are zero} \end{cases}$

Maxwell's equation:

Mech. eq. of motion: $\varrho_m \partial_t^2 u_z = \partial_x T_5 + \partial_y T_4$ $\rightarrow$ $\begin{cases} c(k^2 - \varrho_m\omega^2)u_z + ek^2\phi = 0 \\ ek^2 u_z - \varepsilon k^2\phi = 0 \end{cases}$

$\operatorname{div} D = 0 \rightarrow k^2\phi = 0$

Maxwell's equation: $\operatorname{div} D = 0$

Dispersion relation $k^2 = 0$

Dispersion relation $k^2\left[(\varepsilon c + e^2)k^2 - \varepsilon\varrho_m\omega^2\right] = 0$

Laplace waves	Laplace waves	'Stiffened sound'
$k^2 = 0$	$k^2 = 0$	$k^2 = k_s^2$, with $\dfrac{\omega^2}{k_s^2} = (1 + \dfrac{e^2}{\varepsilon c})\dfrac{c}{\varrho_m}$ (DS)
$D_x = jk_x\varepsilon_0\phi$ $D_y = jk_y\varepsilon_0\phi$ $D_z = 0$	$u_z = 0$	$u_z = \dfrac{\varepsilon\phi}{e}$
	$T_4 = -jk_y e\phi$	$T_4 = -jk_y(1 + \dfrac{e^2}{\varepsilon c})\dfrac{\varepsilon c}{e}\phi$
	$T_5 = -jk_x e\phi$	$T_5 = -jk_x(1 + \dfrac{e^2}{\varepsilon c})\dfrac{\varepsilon c}{e}\phi$
	$D_x = jk_x\varepsilon\phi$ $D_y = jk_y\varepsilon\phi$	$D_x = 0$ $D_y = 0$
(FSL)	(PL)	(PS)

Waves with real positive ω and k_x to be used for superposition

a) The Laplace wave in free space which decays exponentially for $y \to +\infty$:	b) The Laplace wave in the piezoelectric medium which decays exponentially for $y \to -\infty$:	c) 'Stiffened sound', incident wave:	d) 'Stiffened sound', reflected wave:	
$k_{ya} = -jk_x$ ϕ_a $D_{ya} = k_x\varepsilon_0\phi_a$	$k_{yb} = +jk_x$ ϕ_b $T_{4b} = k_x e\phi_b$ $D_{yb} = -k_x\varepsilon\phi_b$	$k_{yc} = +\sqrt{k_s^2 - k_x^2}$ ϕ_c $T_{4c} = -jk_{yc}(1 + \dfrac{e^2}{\varepsilon c})\dfrac{\varepsilon c}{e}\phi_c$ $D_{yc} = 0$	$k_{yd} = -k_{yc}$ ϕ_d $T_{4d} = +jk_{yc}(1 + \dfrac{e^2}{\varepsilon c})\dfrac{\varepsilon c}{e}\phi_d$ $D_{yd} = 0$	(L)

Boundary conditions

$T_{4b} + T_{4c} + T_{4d} = 0$ (B1)

$D_{yb} = D_{ya}$ (B2)

$\phi_b + \phi_c + \phi_d = \phi_a$ (B3)

$\left[k_x + j(1 + \dfrac{\varepsilon}{\varepsilon_0})(\dfrac{\varepsilon c}{e^2} + 1)k_{yc}\right]\phi_c + \left[k_x - j(1 + \dfrac{\varepsilon}{\varepsilon_0})(\dfrac{\varepsilon c}{e^2} + 1)k_{yc}\right]\phi_d = 0$ (A)

Solutions

I. Reflection of 'stiffened sound' at surface:	II. Bleustein-Gulyaev wave	
$k_x^2 < k_s^2 \rightarrow k_{yc}$ real. ϕ_d, ϕ_a, ϕ_b to be calculated from ϕ_c. $|\phi_d| = |\phi_c|$. $k_{yd} = -k_{yc}$.	$k_{yc} = \dfrac{\varepsilon_0}{\varepsilon_0 + \varepsilon}\cdot\dfrac{e^2/\varepsilon c}{1 + e^2/\varepsilon c}jk_x$ $\phi_d = 0$ $k_x^2 + k_{yc}^2 = k_s^2$ $\dfrac{\omega^2}{k_x^2} = \dfrac{c}{\varrho_m}(1 + \dfrac{e^2}{\varepsilon c})\left[1 - (\dfrac{\varepsilon_0}{\varepsilon_0 + \varepsilon}\cdot\dfrac{e^2/\varepsilon c}{1 + e^2/\varepsilon c})^2\right]$	(K) (DBG)

F. C. Eversteyn & B. H. Put: Influence of AsH_3, PH_3, and B_2H_6 on the growth rate and resistivity of polycrystalline silicon films deposited from a SiH_4-H_2 mixture.
J. Electrochem. Soc. **120**, 106-110, 1973 (No. 1).   *E*

C. Foster, W. H. Kool (both with F.O.M.-Instituut voor Atoom- en Molecuulfysica, Amsterdam), W. F. van der Weg (Philips Research Labs., Dept. Amsterdam) & H. E. Roosendaal (F.O.M.-Inst. A. & M., Amsterdam): Random stopping power for protons in silicon.
Radiation Effects **16**, 139-140, 1972 (No. 1/2).

A. J. Fox: Plane-wave theory for the optical grating guide.
Philips Res. Repts. **28**, 306-346, 1973 (No. 4).   *M*

R. G. Gossink: Glasfibers voor optische communicatie.
Klei en Keramiek **23**, 22-28, 1973.   *E*

J. Graf & R. Polaert: Channel image intensifier tubes with proximity focussing. Application to low-light level observation.
Acta Electronica **16**, 11-22, 1973 (No. 1). *(Also in French.)*   *L*

D. Hennings: A study of the incorporation of niobium pentoxide into the perovskite lattice of lead titanate.
Reactivity of solids, Proc. 7th Int. Symp., Bristol 1972, pp. 149-159.   *A*

B. Hill & K. P. Schmidt: New page-composer for holographic data storage.
Appl. Optics **12**, 1193-1198, 1973 (No. 6).   *H*

M. H. H. Höfelt: 'Elastic' constants and wave phenomena in bubble lattices.
J. appl. Phys. **44**, 414-418, 1973 (No. 1).   *E*

H. Hörster, E. Kauer, F. Kettel & A. Rabenau (Max-Planck-Institut für Festkörperforschung, Stuttgart): Analyses de diagrammes et de transformations de phase à hautes températures par mesures électriques.
Etude des transformations cristallines à haute température au-dessus de 2000 K, Coll. Int. C.N.R.S. No. 205, Odeillo 1971, pp. 39-46; 1972.   *A*

B. B. van Iperen & H. Tjassens: An accurate bridge method for impedance measurements of IMPATT diodes.
Microwave J. **15**, No. 11, 29-33, 1972.   *E*

G. A. M. Janssen, J. M. Robertson & M. L. Verheijke: Determination of the composition of thin garnet films by use of radioactive tracer techniques.
Mat. Res. Bull. **8**, 59-64, 1973 (No. 1).   *E*

W. H. de Jeu, J. van der Veen & W. J. A. Goossens: Dependence of the clearing temperature on alkyl chain length in nematic homologous series.
Solid State Comm. **12**, 405-407, 1973 (No. 5).   *E*

N. Kaplan, E. Dormann (both with Technische Hochschule, Darmstadt), K. H. J. Buschow & D. Lebenbaum (Hebrew University, Jerusalem): Magnetic anisotropy and conduction-electron exchange polarization in ferromagnetic (rare-earth)Al_2 compounds.
Phys. Rev. B **7**, 40-49, 1973 (No. 1).   *E*

F. M. Klaassen: Investigation of low level flicker noise in MOS transistors.
Le bruit de fond des composants actifs semi-conducteurs, Coll. Int. C.N.R.S. No. 204, Toulouse 1971, pp. 111-113; 1972.   *E*

W. F. Knippenberg, G. Verspui & G. A. Bootsma: Phases of silicon carbide.
Etude des transformations cristallines à haute température au-dessus de 2000 K, Coll. Int. C.N.R.S. No. 205, Odeillo 1971, pp. 163-170; 1972.   *E*

R. Koppe: Automatische Abbildung eines planaren Graphen in einen ebenen Streckengraphen.
Computing **10**, 317-333, 1972 (No. 4).   *H*

J. Krüger & W. Jasmer: Ein neues elektrooptisches Verfahren zur Erzeugung von Mikromustern.
Mikroelektronik 5 (Kongreß INEA, München 1972), 202-213, 1973.   *H*

J.-P. Krumme, G. Bartels & W. Tolksdorf: Magnetic properties of liquid-phase epitaxial films of $Y_{3-x}Gd_xFe_{5-y}Ga_yO_{12}$ for optical memory applications.
Phys. Stat. sol. (a) **17**, 175-179, 1973 (No. 1).   *H*

P. K. Larsen & A. B. Voermans: Origin of the conductivity minimum and the negative magnetoresistance in n-type sulpho-spinels.
J. Phys. Chem. Solids **34**, 645-650, 1973 (No. 4).   *E*

P. R. Locher: Nuclear magnetic resonance of ^{59}Co on A and B sites in Co_3S_4.
Physics Letters **42A**, 490, 1973 (No. 7).   *E*

H. H. van Mal: A $LaNi_5$-hydride thermal absorption compressor for a hydrogen refrigerator.
Chemie-Ing.-Technik **45**, 80-83, 1973 (No. 2).   *E*

J. R. Mansell: A study of an experimental TV pick-up tube incorporating a channel plate.
Acta Electronica **16**, 113-122, 1973 (No. 1). *(Also in French.)*   *M*

I. C. P. Millar, D. Washington & D. L. Lamport: Channel electron multiplier plates in X-ray image intensification.
Adv. in Electronics & Electron Phys. **33A**, 153-165, 1972.   *M*

A. Mircea, J. Magarshack & A. Roussel: Etude du bruit basse fréquence des diodes Gunn au GaAs et de sa corrélation avec le bruit de modulation de fréquence des oscillateurs à diode Gunn.
Le bruit de fond des composants actifs semi-conducteurs, Coll. Int. C.N.R.S. No. 204, Toulouse 1971, pp. 217-223; 1972.   *L*

D. Muilwijk (Philips' Telecommunicatie Industrie B.V., Hilversum): Comparison and optimization of multiplexing and modulation methods for a group of radio networks.
Philips Res. Repts. **28**, 347-376, 1973 (No. 4).

A. G. van Nie: An operator treatment of modulated carriers.
Proc. IEEE **61**, 131-132, 1973 (No. 1).   *E*

S. G. Nooteboom & I. H. Slis (Institute for Perception Research, Eindhoven): The phonetic feature of vowel length in Dutch.
Language and Speech **15**, 301-316, 1972 (No. 4).

A. van Oostrom: Field ion microscopy.
Acta Electronica **16**, 59-71, 1973 (No. 1). *(Also in French.)*   *E*

R. A. Ormerod & R. W. A. Gill: Electroforming for electronics.
Trans. Inst. Met. Finish. **51**, 23-26, 1973 (No. 1).   *M*

J. A. Pals: Measurements of the surface quantization in silicon *n*- and *p*-type inversion layers at temperatures above 25 K.
Phys. Rev. B **7**, 754-760, 1973 (No. 2).   *E*

R. I. Pedroso & G. A. Domoto (Columbia University, New York): Inward spherical solidification — solution by the method of strained coordinates.
Int. J. Heat & Mass Transfer **16**, 1037-1043, 1973 (No. 5).   *N*

A. Rabenau (Max-Planck-Institut für Festkörperforschung, Stuttgart) **& H. Rau:** Über die Systeme Te-TeCl$_4$ und Te-TeBr$_4$.
Z. anorg. allg. Chemie **395**, 273-279, 1973 (No. 2/3).   *A*

H. Rau & J. F. R. Guedes de Carvalho: Equilibria of the reduction of NiO and CoO with hydrogen measured with a palladium membrane.
J. chem. Thermodyn. **5**, 387-391, 1973 (No. 3).   *A*

W. Rey, R. Bernard (Hôpital St. Pierre, Bruxelles) **& H. Vainsel** (Hôp. St. Pierre, Br.): Adaptivité en surveillance de l'ECG et détection par ordinateur des arythmies.
J. Inform. méd. IRIA, mars 1973, **1**, 185-191.   *B*

J. M. Robertson, M. J. G. van Hout, M. M. Janssen & W. T. Stacy: Garnet substrate preparation by homoepitaxy.
J. Crystal Growth **18**, 294-296, 1973 (No. 3).   *E*

C. J. M. Rooymans: Growth and applications of single crystals of magnetic oxides.
Etude des transformations cristallines à haute température au-dessus de 2000 K, Coll. Int. C.N.R.S. No. 205, Odeillo 1971, pp. 151-162; 1972.   *E*

A. Roussel & A. Mircea: Identification of a bivalent flaw in *n*-GaAs from noise measurements.
Solid State Comm. **12**, 39-42, 1973 (No. 1).   *L*

G. Salmer, J. Pribetich (both with Université des Sciences et Techniques de Lille), **A. Farrayre & B. Kramer:** Theoretical and experimental study of GaAs IMPATT oscillator efficiency.
J. appl. Phys. **44**, 314-324, 1973 (No. 1).   *L*

J. Schramm & K. Witter: Gas discharges in very small gaps in relation to electrography.
Appl. Phys. **1**, 331-337, 1973 (No. 6).   *H*

L. A. Æ. Sluyterman & J. Wijdenes: Benzoylamidoacetonitrile as an inhibitor of papain.
Biochim. biophys. Acta **302**, 95-101, 1973 (No. 1).   *E*

M. J. Sparnaay: Surface analysis with the aid of low-energy electrons. The case of semiconductor surfaces compared with other methods.
J. radioanal. Chem. **12**, 101-114, 1972 (No. 1).   *E*

P. J. Strijkert, R. Loppes (University of Liège) **& J. S. Sussenbach:** The actual biochemical block in the *arg-2* mutant of *Chlamydomonas reinhardi*.
Biochem. Genet. **8**, 239-248, 1973 (No. 3).   *E*

A. L. Stuijts: Sintering theories and industrial practice.
Materials Science Research **6**: Sintering and related phenomena, editor G. C. Kuczynski, pp. 331-350, 1973.   *E*

J. L. Teszner (Ecole Normale Supérieure, Paris) **& D. Boccon-Gibod:** Dependence of Gunn threshold on transverse magnetic field in coplanar epitaxial devices.
Phys. Stat. sol. (a) **15**, K 11-14, 1973 (No. 1).   *L*

A. Thayse & M. Davio: Boolean differential calculus and its application to switching theory.
IEEE Trans. **C-22**, 409-420, 1973 (No. 4).   *B*

J. B. Theeten, L. Dobrzynski (Institut Supérieur d'Electronique du Nord, Lille) **& J. L. Domange** (E.N.S.C.P., Paris): Vibrational properties of the adsorbed monolayer on face-centered cubic crystals.
Surface Sci. **34**, 145-155, 1973 (No. 1).   *L*

J. F. Verwey, J. H. Aalberts & B. J. de Maagt: Drift of the breakdown voltage in highly doped planar junctions.
Microelectronics and Reliability **12**, 51-56, 1973 (No. 1).   *E*

J. Visser: The gas composition in r.f. sputtering systems and its dependency on the pumping method.
Journées de Technologie du Vide, Versailles 1972 (Suppl. Le Vide No. 157), pp. 51-66.   *E*

M. T. Vlaardingerbroek & Th. G. van de Roer (Eindhoven University of Technology): On the theory of punch-through diodes.
Appl. Phys. Letters **22**, 146-148, 1973 (No. 4).   *E*

J. A. Weaver: Optical character recognition.
Physics Bull. **24**, 277-278, 1973 (May).   *M*

J. S. C. Wessels, O. van Alphen-van Waveren & G. Voorn: Isolation and properties of particles containing the reaction-center complex of Photosystem II from spinach chloroplasts.
Biochim. biophys. Acta **292**, 741-752, 1973 (No. 3).   *E*

M. V. Whelan: Resistive MOS-gated diode light sensor.
Solid-State Electronics **16**, 161-171, 1973 (No. 2).   *E*

H. W. de Wijn (State University of Utrecht), **A. M. van Diepen & K. H. J. Buschow:** Effect of crystal fields on the magnetic properties of samarium intermetallic compounds.
Phys. Rev. B **7**, 524-533, 1973 (No. 1).   *E*

A. W. Woodhead & G. Eschard: Introduction *(to Channel electron multipliers issue of Acta Electronica)*.
Acta Electronica **16**, 8 (in English), 9 (in French), 1973 (No. 1).   *M, L*