PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 1

The onboard computer of the Netherlands astronomical satellite (ANS)

G. J. A. Arink

This third article on the Netherlands astronomical satellite (ANS) deals with the digital computer on board the satellite. The orbit that the satellite will describe is such that communication between the ground station and the satellite will be possible only once in every 12 hours. This means that upon each ground contact the measurement program for the next period of 12 hours must be transmitted to the satellite, and during this period the results of the measurements have to be stored in order to be sent back to the ground station when the next contact is made. The computer stores all this data and plays a central part in the attitude control, in carrying out the measurement program and in the communication with the ground station. In designing the satellite it was of course necessary to make the maximum use of the space available while keeping weight and energy consumption to a minimum. Faced with these limitations the computer designers succeeded in producing an onboard computer that has a storage capacity of 0.46×10^6 bits yet weighs only 8 kg, occupies no more than 10 dm^3 and runs on a power of less than 8 watts.

Introduction

In 1974 a Scout launch vehicle will be launched from an American launching site to put into orbit around the Earth a small satellite made in the Netherlands for the purpose of carrying out astronomical research [1]. Known as ANS (standing for *Astronomische Nederlandse Satelliet*), the satellite carries three astrophysical observation systems: an ultraviolet spectrometer and two systems for measuring X-radiation [2]. These instruments are rigidly connected with the frame of the satellite; when an instrument has to be pointed at a celestial object for observation, this is done by turning the whole satellite into the desired direction and keeping it pointed at the object during the measurement [3]. The various operations of the satellite are coordinated and controlled by a small digital computer on board the satellite. This onboard computer performs the calculations for the attitude-control system during manoeuvring and alignment with an object, sends commands to the observation and other systems of the satellite, and collects, processes and stores the data generated by the astronomical instruments and data on the operation of the satellite.

In the main there will only be one ground station for the satellite, the ESRO station at Redu in Belgium. Since the satellite will describe a virtually polar orbit, communication with the satellite will be possible from this ground station only once in every 12 hours. The computer must therefore have sufficient storage capacity to carry the command instructions for attitude control and operation of the measurement systems

Ir G. J. A. Arink, M.Sc., Mech. Eng., is with Philips Research Laboratories, Eindhoven. Ir Arink is the Subsystem Manager for the onboard computer in the ANS project.

[1] W. Bloemendal and C. Kramer, The Netherlands astronomical satellite (ANS), Philips tech. Rev. **33**, 117-129, 1973 (No. 5).

[2] A forthcoming issue of this journal will contain two articles on the scientific experiments carried out with ANS. They are also described in: The scientific mission and the experiments of the Astronomical Netherlands Satellite, Report RP 72/1, June 1972, Nederlands Instituut voor Vliegtuigontwikkeling en Ruimtevaart (NIVR), Delft.

[3] P. van Otterloo, Attitude control for the Netherlands astronomical satellite (ANS), Philips tech. Rev. **33**, 162-176, 1973 (No. 6).

during each interval of 12 hours, as well as storage capacity for the results of the observations made during each interval. At the next ground contact these results are then transmitted to the ground station, while new instructions are sent to the satellite for storage in the onboard computer.

The choice of the onboard computer

In unmanned space vehicles for scientific research the memory devices at first used for storing operating instructions and measurement results were tape recorders. These devices are subject to failure, however, because of tape wear or defective moving parts. It was therefore decided that the ANS onboard computer should have a type of memory without moving parts, i.e. a core-store type. It was also desired to make calculations for attitude control and data reduction of the measurement results, to allow a low signal rate to be used during ground contacts. This led to the decision to build a small digital computer into the satellite.

The presence of a computer in the satellite offers many advantages, both for testing during the assembly of the satellite and for preparation for the launch. The onboard computer can also be used immediately after the launch for testing the satellite equipment, enabling performance measurements to be made that would be difficult to carry out on Earth. Another advantage is that changes can be made in the measurement and control programs after launching, so as to adapt the objectives of the project to new scientific discoveries, made for example with other satellites.

To carry out these various tasks it was found that the ANS needed a computer with a storage capacity of 4k (4096) words of 16 bits to store the measurement and control programs and the operational program, the *program memory*, and a memory with a capacity of 24k words of 16 bits for data storage, the *data memory*. Out of the total power available in the satellite only 8 W was available for the computer. At the beginning of the project none of the existing aerospace computers met this combination of requirements [4]. A particular difficulty was the power-consumption requirement. By way of illustration *Table I* shows some figures characteristic of the state of the art in 1969 side by side with the values required (and later achieved) for the onboard computer of ANS.

Table I. Comparison of the ANS onboard-computer specifications with the 'state of the art' in 1969.

	Typical values in 1969	Values specified for ANS computer
Memory capacity (in 16-bit words)	4k-32k	28k
Weight (kg)	10-25	max. 8
Volume (litres)	6-20	max. 10
Power (watts)	50-200	max. 8
Total time for an addition (microseconds)	4-9	128

Before going any further into the operation and design of the computer, a brief description will first be given of the main parts of the computer and the various tasks which the computer is required to carry out in combination with the other systems of the satellite.

The general design of the computer

The general layout of the computer developed for ANS does not differ from the normal layout of a digital computer (see *fig. 1*). Like other computers, it has an arithmetic and logic unit (*ALU*), a control unit (*CU*) which ensures that the different operations take place in the proper sequence, an input-output control (*IOC*),

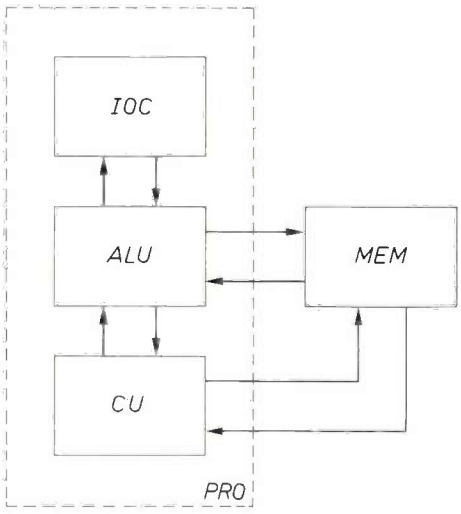


Fig. 1. General layout of the onboard computer. *ALU* arithmetic and logic unit. *CU* control unit. *IOC* input-output control. *MEM* memory. The first three units are collectively referred to as the processor *PRO*.

Fig. 2. Construction of the onboard computer. The module with the connectors contains the processor, the power-supply circuits, the block-decoding and address-decoding circuits, and the timing circuit of the memory. The seven other modules each contain a memory block with associated address-selection circuits. (Dimensions 13×20×44 cm.)

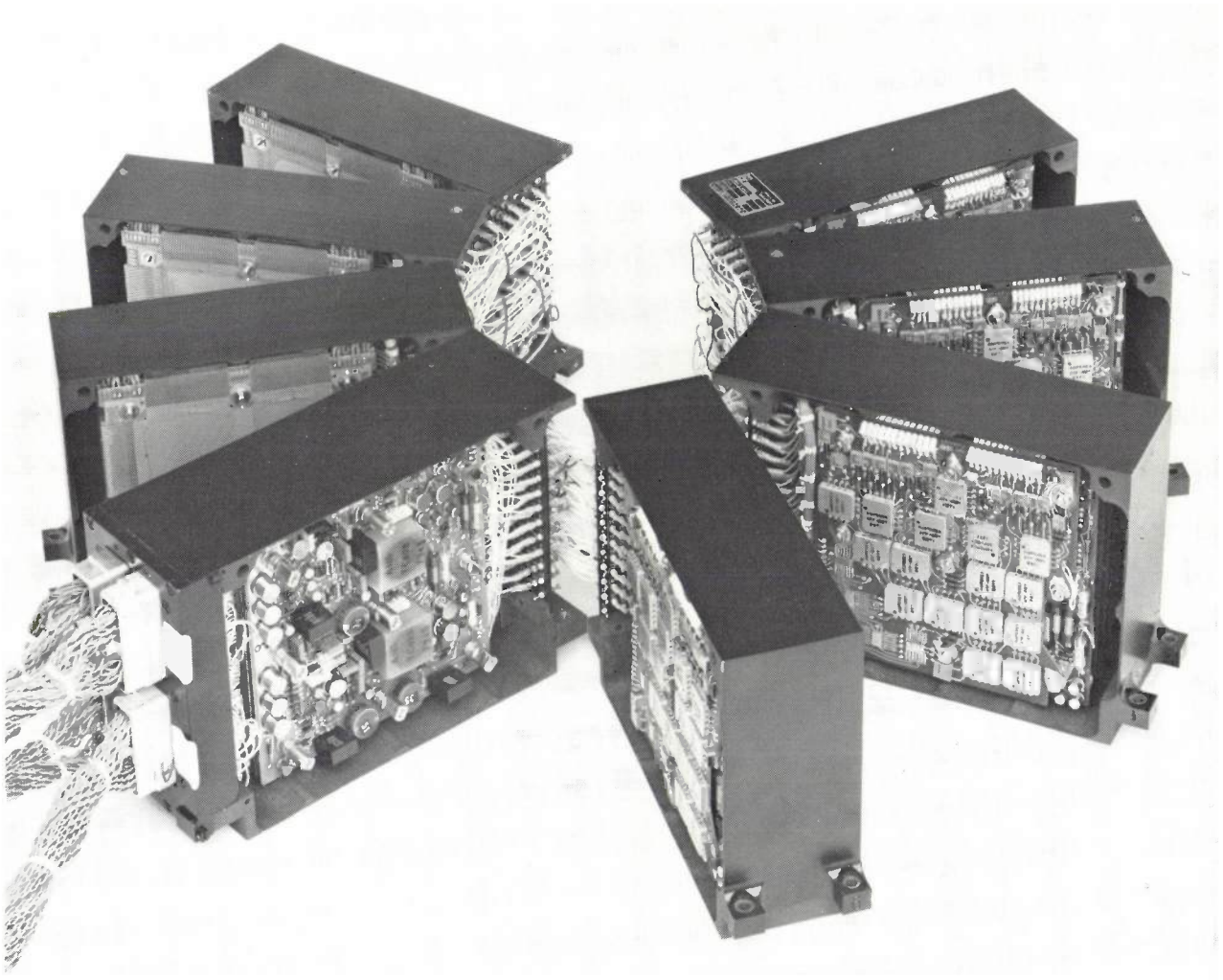


Fig. 3. The computer opened up. Since the interconnections between the modules are along one side of the blocks, it is possible to open the computer as shown without breaking any connections (all the connections are soldered). It is also possible to remove the metal module cases without dismantling the electronic units, and to 'unfold' the assembly of printed-circuit boards. Parts of the core matrices can be seen in the three memory blocks on the left.

which controls the interfacing with the communication, measurement and attitude control systems, and finally a memory (*MEM*).

The memory is a core-store type consisting of seven identical blocks each with a capacity of 4096 words of 16 bits. The total storage capacity is thus 0.46×10^6 bits. By means of a command from the ground station one of these blocks is assigned the function of program memory (*PMY*), while the other six act as the data memory (*DMY*). This implies that if a defect should occur in one of the memory blocks the essential tasks of the onboard computer can still be carried out.

The control unit contains a clock-pulse generator with two 8 MHz quartz-crystal oscillators (one as a standby). This generator delivers clock pulses of different frequencies, down to $\frac{1}{8}$ Hz, not only for the computer, but also for the other subsystems of the satellite.

The arithmetic and logic unit, the control unit and the input-output control are together referred to as the processor (*PRO*). In the construction of the computer the processor is accommodated together with the power-supply circuits, the block-decoder and address-decoder circuits and the timing unit of the memory in one module (see *figs. 2* and *3*). The other seven modules each contain a memory block with the associated address-selector circuits.

Since a high computing speed is not needed, serial transfer could be adopted nearly everywhere in the computer (all numerical operations and transfers are made bit by bit), which requires fewer circuits than for parallel transfer. The low energy consumption of the computer is mainly achieved by switching on the

[4] D. O. Baechler, Trends in aerospace digital computer design, IEEE Computer Group News 2, No. 7, 18-23, Jan. 1969.

+5 volt and —5 volt supplies to an individual memory block only for as long as information is to be read in or read out of it. Low-power TTL logic (transistor-transistor logic) are used for most of the logic circuits. To keep the weight down, light materials such as magnesium and titanium were used in the construction of the computer and its housing.

The signal arriving from the ground station consists of messages of 32 bits, transmitted at an information rate of 400 bits per second. The 32 bits include four parity bits — check bits to safeguard against transmission errors — and the other bits contain the information. The messages fall into two groups: messages conveying a command to be carried out immediately

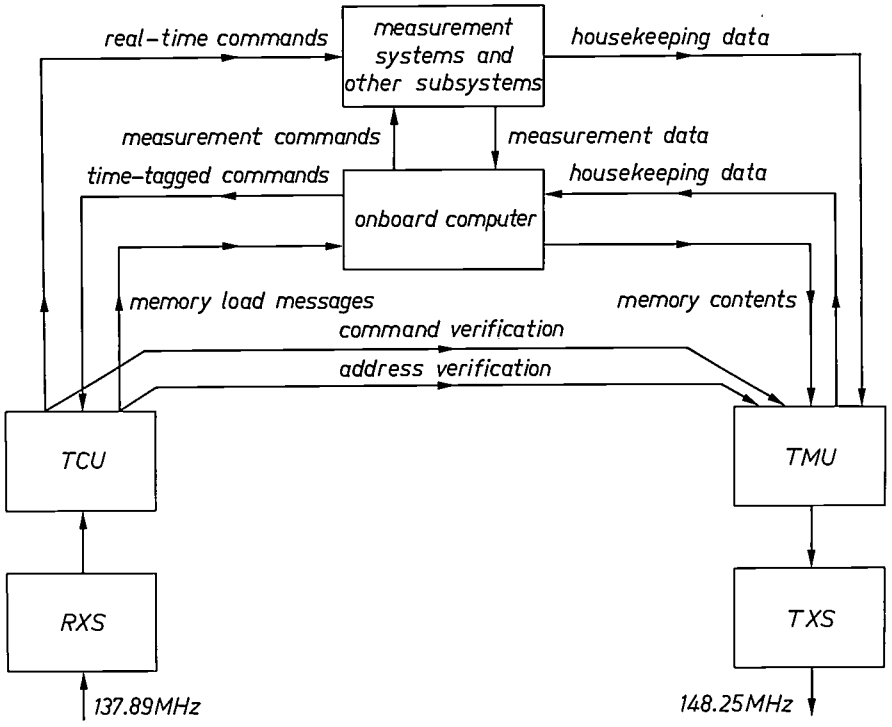


Fig. 4. Diagram illustrating the telecommunication system of the satellite and its interfacing with the other systems. The telecommand unit (*TCU*) decodes the information arriving via the receivers *RXS* into commands to be directly executed and messages to be stored in the memory. These messages can consist of new software instructions, parameters and commands to be executed later. The telemetry unit is a coding unit which collects the housekeeping data, a selection of data on the operation of the satellite, which are temporarily stored in the memory. During ground contact the results of observations during the previous period of 12 hours are combined with the housekeeping data and with data for verifying the correct reception of the new information by *TCU* to form a signal that is transmitted at high speed (4096 bits/s) by the transmitter *TXS* (high-speed frame). In the interim periods a low-speed frame is transmitted (speed 128 bits/s) for tracking the satellite. This frame contains the housekeeping data only.

Interfacing with the communication system

For the exchange of command instructions and measurement results between satellite and ground station the satellite carries a pulse-code-modulation (PCM) telecommunication system (*fig. 4*). At the receiving end it consists of two receivers *RXS*, which operate in the 148.25 MHz band, and a decoding telecommand unit *TCU*; at the transmitting end it consists of a coding telemetry unit *TMU* and a transmitter *TXS*, which operates at 137.89 MHz. The figure also shows the various flows of information between these units, the onboard computer and the measurement systems. (real-time commands) and messages conveying information to be stored in the computer memory (memory load messages). The real-time commands are operating instructions to switch instruments on and off, to operate shutters in observation systems, etc.; they are fed direct to the appropriate system, which immediately carries out the command.

The memory load messages are passed to the computer. They consist of a memory address of 12 bits, an information word of 16 bits for storage at this address and four parity bits. The word may be a part of the 'satellite-operation program' *SOP*, or an instruction

from a new program for the computer. *SOP* contains all the activities of the satellite for the next period of 12 hours. It is a list giving the required commands, set-points for the attitude control and other data for each activity, provided with a 'time tag', a word that specifies accurately in seconds when the activity is to be carried out. This exact instant in time is expressed in 'satellite-calendar time', indicated by the contents of a counting register that receives a pulse from the clock-pulse generator every second (this register is located in the coding telemetry unit *TMU*). The commands in *SOP* can have the same function as the real-time commands, or they may be commands to bring the scientific instruments into a particular mode of operation (measurement commands). The decoding and implementation of *SOP* will be dealt with later in this article when we come to the subject of software.

Together with the stored results of the measurements it is also necessary to transmit to the ground station the 'housekeeping' information, a selection of data on the state of the observation systems and of the other equipment in the satellite (digitized values of voltages, temperatures, gas pressures, etc.). The housekeeping data is collected by *TMU* and coded into a signal referred to as the low-speed telemetry frame, which consists of groups of 128 words of 8 bits. A complete group of housekeeping data is made available in 8 seconds, so that the information rate is 16 words or 128 bits per second. While the satellite is out of sight of the ground station it transmits this signal with a power of 0.12 W. It is regularly used for determining the orbit of the satellite with the aid of the existing network of satellite-tracking stations on the ground. During normal operation of the satellite the information content of the signal is not used, but in certain emergency situations it may be necessary to collect the information via these stations.

The housekeeping data is continuously read into the onboard computer and stored for 8 seconds in the program memory. A small selection is stored in a block of the data memory and kept there until the next contact with the ground station. During ground contact the housekeeping data and also the measurement data and the observation program for the previous 12 hours are transmitted at a rate of 4096 bits per second with a power of 1.2 W. This is called the high-speed telemetry frame. In this way the entire contents of the memory are transmitted to the ground station in less than 2 minutes, a process known as 'dumping'. During this transmission the low-speed telemetry information is interleaved in the high-speed frame.

The three input and output activities of the computer described above — loading the memory, dumping, and the storage of housekeeping data — take priority over all other activities of the computer. Operations controlled by the onboard software are interrupted for these 'cycle-stealing' operations if they require access to the memory. This is done by means of a special switching circuit in the control unit of the computer, which is commanded by *TCU*, *TMU* and the clock-pulse generator.

Interfacing with the attitude-control system

For each observation the satellite has to be oriented in a particular attitude. A coordinate system is used which is rigidly connected with the satellite (see *fig. 5*). For determining the attitude of the satellite a number of sensors are available: a magnetometer (*MGM*), six coarse solar sensors (*CSS*) and an intermediate solar sensor (*ISS*) for pointing the z-axis at the Sun; two fine solar sensors (*FSSX* and *Y*), a horizon sensor (*HSE*) and star sensor (*SSE*) for pointing the x-axis at an object. The actuators used for rotating the satellite about the three axes are three magnetic torquing coils (*MCX*, *Y* and *Z*) and three reaction wheels (*RWX*, *Y* and *Z*).

The actuators are controlled by the onboard computer and by the 'attitude-control logic' (*ACL*), which is a 'wired-logic' circuit. During the coarse and intermediate operating modes of the attitude-control system,

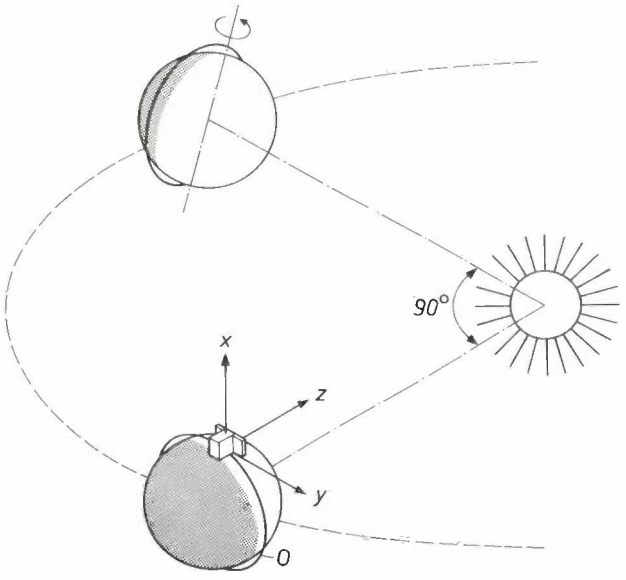


Fig. 5. Orientation of the orbital plane *O* of the satellite at two times separated by three months. The orbital plane is chosen in such a way that it rotates with the Sun, thus keeping the satellite approximately over the terminator and permanently outside the cone of shadow of the Earth [5]. The z-axis of the coordinate system is kept continuously pointed at the Sun. The instruments measure radiation incident along the x-axis. The rotation of the satellite about the z-axis enables the instruments to observe objects along a circular arc in space. Because of the rotation of the orbital plane this means that the whole celestial sphere can be observed in six months.

[5] This is explained in the article of note [1], page 122.

which serve mainly for reducing the initial spin (detumbling) and for pointing the z-axis at the Sun, only the attitude-control logic is in operation; in the fine-pointing mode, during which the instruments are pointed at a celestial object for observation, the control is taken over by the onboard computer. For this purpose the computer receives the output signals from the horizon sensor, the star sensor, the fine solar sensors and from the subsystem used for measuring hard X-radiation. The computer uses the sensor output signals to calculate the torque setpoint values for the reaction wheels at one-second intervals from one of the six algorithms that apply for the different submodes of the fine pointing [6].

Interfacing with the observation and other systems

The third main task of the onboard computer is connected with the operation of the three observation systems. In each of these systems there are a number of modes of operation, which are related to the submode of the attitude control of the satellite, the time resolution and energy resolution of the measurements, or whether or not a shutter is open or closed (e.g. in connection with photomultiplier dark-current measurements). It may also relate to calibration measurements using calibration sources present in the satellite [7]. These measurement modes are initiated at the required moment by a command from the satellite-operation program *SOP*. As described above, the systems are switched on and off by means of operating commands received directly from the ground station, or through the onboard software if they are contained in the *SOP*.

The other systems in which the onboard computer is involved include the pulse generator *PYE*, which sets various mechanisms in operation at certain moments after the launching (e.g. the yo-yo actuator that stops the initial spin of the satellite). The same circuit initiates the power supply to the computer shortly after deployment of the solar panels, and gives the 'set busy' signal for starting the computer 5 seconds later.

Finally the computer has a separate output channel through which, shortly before the launching, all the information present in the memory can be dumped to the test station.

Operation of the computer

Fig. 6 shows a simplified block diagram of the onboard computer, indicating only the units that fulfil a function in executing software instructions. The three direct memory activities not controlled by software but by hardware (loading, dumping and storage of housekeeping data) will be discussed later. The transfer of data in the computer is mainly done by serial shifting, the word length is 16 bits and the read-out of the data is in the fixed-point notation. The two's-complement notation was chosen for negative numbers [8]; the most significant bit of each number (the first) is the sign bit (0 is positive, 1 is negative).

The arithmetic and logic unit (*ALU*) contains three 16-bit registers *A*, *M* and *D*, and a serial adder. The flow of data to and from the memory goes via the *M* register. The serial adder sums the contents of the *A* and *M* registers and transfers the result to the *D* register. From here the information can be passed for a further operation to the *A* or the *M* register. Apart from through the *M* register, data can also enter and leave the arithmetic and logic unit through the *D* register; this is used for the exchange of information with the other systems of the satellite.

The control unit *CU* contains a number of 12-bit registers with which it controls the processing of an instruction. Such an instruction cycle comprises two phases, the *fetch phase* and the *execute phase*, each of which consists in its turn of a number of steps. During the fetch phase the instruction is taken from the memory. The address where the instruction is stored, and which is located in the *operation register*, is first transferred to the *address register*. This register is connected with the address decoder in the control unit of the memory. The instruction is now fetched from the memory and read into the *M* register of the arithmetic and logic unit. The instruction is then immediately read back into the same address in the memory, since on read-out from a core store the information is destroyed. The choice of the memory block is made with the aid of the *block-selection register*; this will be dealt with later, together with the direct input and output (cycle-stealing) activities.

The instruction cycle continues as follows. The address part of the instruction, which gives the address of the relevant word (12 least-significant bits), is transferred from the *M* register to the address register, and the execution part (12 most-significant bits) is transferred to the *instruction register*. In the execute phase the contents of the address specified are then transferred from the memory to the *M* register. The *sequencer* now decodes from the state of the instruction register the operation to be executed, and at the end of the cycle delivers control signals which cause this operation to be carried out. In the course of the cycle the contents of the operation register are incremented by 1, so that the next instruction from the memory can now be fetched (the software instructions are generally located at successive addresses in the program memory). By means of a *jump instruction* it is also possible to read a new address into the operation register, so that the next instruction need not necessarily always have to

be fetched from the next address. The jump instruction may or may not be conditional (see below).

The memory *MEM* contains, in addition to the seven memory blocks, a control unit referred to as the memory distribution unit (not to be confused with the control unit *CU* of the computer), which comprises block-decoding circuits, address-decoding circuits, a timing unit and a power supply and distribution unit.

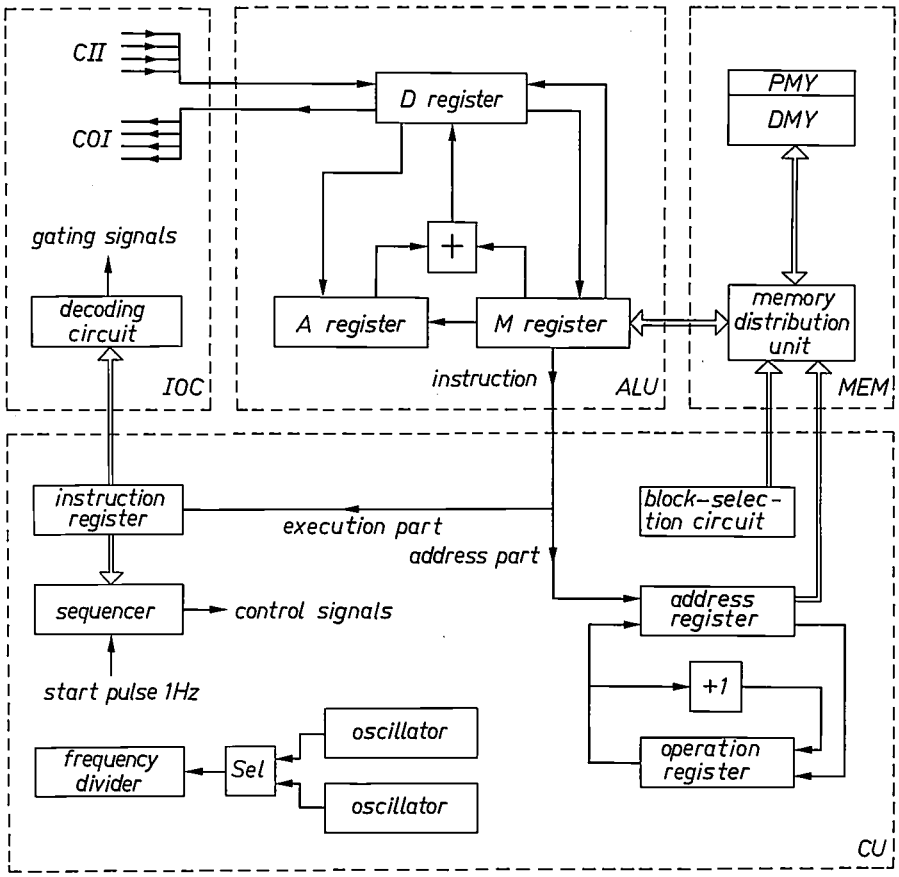


Fig. 6. Simplified block diagram of the onboard computer. The various sections of the arithmetic and logic unit *ALU*, the control unit *CU*, the memory *MEM* and the input-output control *IOC*, are connected by lines indicating the information flow. Single lines refer to serial transfer and double lines to parallel transfer.

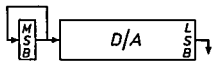
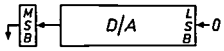
The operation of the memory will be discussed under a separate heading.

The computer works with a total of 21 instructions. *Table II* gives a list of these instructions, using a mnemonic code notation, and indicates the various functions by means of symbols (the term 'flag bit of *BSR/DA*' will be discussed when we come to the subject of memory-block selection). The instructions can now be classified by function into the following categories: load and store instructions, for reading data from the memory into a register or vice versa; arithmetic instructions, shift instructions, control instructions and input/output instructions. There are always two versions, *P* and *N*, of the instructions involving the processing of a word; in the *N* version the negation of the word is processed. The instructions can also be distinguished according to the method of coding; see *Table III*. There are then four types: memory-reference instructions *PMR* and *DMR*, for which a word has to be fetched from the program memory or the data memory (these instructions here are all of the one-address type); generic instructions *G*, which are general instructions for which only four bits are needed; input/output instructions *IO*; and shift instructions *SH*.

In the control unit of the computer the most important function is carried out by the sequencer. This controls not only the arithmetic operations but also delivers the signals controlling the various stages of the fetch

[6] See Table II in the article of note [3].
[7] See the articles quoted in note [2].
[8] In this notation each bit of a word is inverted, and 1 is added to the result. For example, the negative version of the binary number 00101 (5) is 11010 + 1 = 11011. A separate sign bit is used to indicate whether the number is positive or negative.

Table II. List of instructions for onboard computer

Category	Code	Function	Type
Load and store	ODP	$(D) \rightarrow [EAP]$	*PMR*
	ODN	$(\bar{D}) \rightarrow [EAP]$	
	DDP	$(D) \rightarrow [EAD]$	*DMR*
	DDN	$(\bar{D}) \rightarrow [EAD]$	
	IAP	$[EAP] \rightarrow (A)$ } if flag bit of	*PMR*
	IAN	$[\overline{EAP}] \rightarrow (A)$ } $BSR/DA = 0$	
	IAP	$[EAD] \rightarrow (A)$ } if flag bit of	*DMR*
	IAN	$[\overline{EAD}] \rightarrow (A)$ } $BSR/DA = 1$	
	IDP	$[EAP] \rightarrow (D)$	*PMR*
	IDN	$[\overline{EAP}] \rightarrow (D)$	
Arithmetic	IMP	$[EAP] + (A) \rightarrow (D)$	*PMR*
	IMN	$[\overline{EAP}] + (A) \rightarrow (D)$	
Shift	SRD		*SH*
	SRA		
	SLD		*SH*
	SLA		
Control	DAP	$(D) \rightarrow (A)$	*G*
	DAN	$(\bar{D}) \rightarrow (A)$	
	STP	set computer in 'soft rest' state until next 1 Hz pulse	*G*
	JUM	$[EAP] \rightarrow (OT)$	*PMR*
	IFN	execute JUM instruction if $(D_1)=1$, otherwise take next instruction	
Input/output	IPT	$(CII) \rightarrow (D)$	*IO*
	OPT	$(D) \rightarrow (COI)$	*IO*

Explanation of symbols	
A	*A* register
D	*D* register
D_1	sign bit of *D* register
OT	operation register
EAP	effective address in program memory
EAD	effective address in data memory
flag bit *BRS/DA*, see memory selection	
CII	computer input interface register
COI	computer output interface register
PMR	program-memory reference
DMR	data-memory reference
G	generic instruction
IO	input/output instruction
SH	shift instruction
+	addition
→	replaces
⏋	discarded
( )	contents of hardware register
[ ]	contents of memory location
MSB	most-significant bit (sign bit)
LSB	least-significant bit

and execution phases. For this purpose the control circuit contains an 8-bit register, called the execution-stage register (*ESR*) and a counter circuit, called the clock-pulse counter (*CPC*), which receives clock pulses at a frequency of 524 288 Hz (see *fig. 7*). At the beginning of the instruction cycle, *ESR* is set from its rest state 11111111 into the state 01111111. When the clock-pulse counter has received a certain number of clock pulses, the next clock pulse enables *ESR* to shift to the state 10111111. At the same time *CPC* is reset to zero and then begins to count the next step, after which *ESR* shifts to the state 11011111, and so on. During each of the nine steps through which the *ESR* is shifted in this way, a logic circuit that decodes the state of the *ESR* delivers a control signal for executing a step of the instruction cycle (e.g. by opening a gate that permits transfer between two registers). The control signals for the last step of this cycle, the arithmetic operation on the word, are delivered by a logic circuit that decodes the state of the instruction register (see fig. 6).

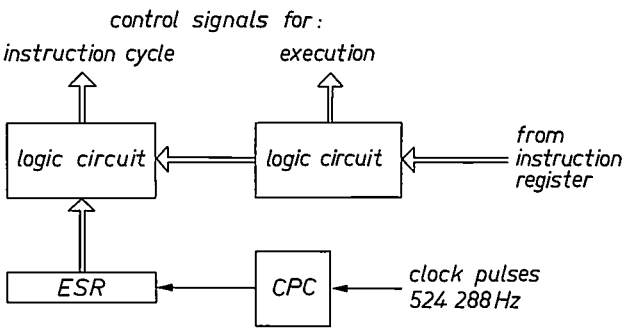


Fig. 7. Block diagram of the section of the sequencer of the control unit that controls the instruction cycle; *ESR* execution-stage register. *CPC* clock-pulse counter.

At 1-second intervals the sequencer receives a starting pulse from the clock-pulse generator, thus starting the part of the program prepared for that particular period, which is terminated with a stop instruction. The time between the execution of a stop instruction and the next 1-Hz pulse is called the soft-rest state. During this period the execution-stage register *ESR* stands by in its rest state, so that a possible request for a direct input or output activity ('cycle stealing') can be executed without delay.

Table III. Coding of instructions. Cross-hatching through a number of bits indicates that they are irrelevant for the instruction concerned. The explanation of the symbols is given in Table II.

instruction ↓	bits of the code word 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	type ↓
IFN	0	0	0	0													PMR
JUM	0	0	0	1													PMR
IAP	0	0	1	0													DMR
IAN	0	0	1	1													DMR
IMP	0	1	0	0	12 bit addresses for program or data memory												PMR
IMN	0	1	0	1													PMR
IDP	0	1	1	0													PMR
IDN	0	1	1	1													PMR
ODP	1	0	0	0													PMR
ODN	1	0	0	1													PMR
DDP	1	0	1	0													DMR
DDN	1	0	1	1													DMR
DAP	1	1	0	0													G
DAN	1	1	0	1													G
IPT	1	1	1	0	0	0	4 bit labels of I/O registers										IO
OPT	1	1	1	0	0	1											IO
SRD	1	1	1	0	1	0	0	no. of steps (0 to 15)									SH
SRA	1	1	1	0	1	0	1										SH
SLD	1	1	1	0	1	1	0										SH
SLA	1	1	1	0	1	1	1										SH
STP	1	1	1	1													G

The control unit also includes the clock-pulse generator. There are two quartz-crystal oscillators operating at a frequency of $2^{23} = 8\,388\,608$ Hz. The stability of these oscillators in a temperature range from 5 to 45 °C is about 10^{-7}, a result achieved by means of an optimized compensation network of resistors and thermistors [9]. At the beginning of the mission a selector switch selects the output of one of the oscillators, while the other stands by ready to take over immediately in the event of a fault. The change-over is effected by the selector switch as soon as the amplitude of the clock signal falls below a critical threshold value. The oscillator signal is fed to a frequency-divider chain which reduces the frequency in 26 steps to $\frac{1}{8}$ Hz. The signal at the frequency 131 072 Hz is used as the master frequency for a frequency-divider in the telemetry-coding unit *TMU*, which generates the satellite-calendar time.

The onboard computer is either in the 'busy' or in the 'idle' state. In the busy state the computer can perform all its activities, while in the idle state it is only capable of generating clock signals. The idle state is the initial state reached after the power has been switched on (and after the return of the power after a short interruption), or after receipt of a 'set-idle' command from the telecommand unit *TCU*. The 'busy' state is entered after receipt of a 'set busy' command from the pulse generator *PYE* shortly after the launch, or upon receipt of a program-allocation command from *TCU*.

Input and output of information to and from the observation systems and the attitude-control system is effected through registers built into these units called the computer-input and computer-output interfaces (*CII* and *COI*), which are connected with the *D* register. During the execution of an *I/O* instruction the first 12 bits are read into the instruction register in the same way as described above. A special decoding circuit in the *I/O* unit now decodes the state of the bits 9 to 12, which indicate which register is to be used (see Table III), and delivers gating signals that connect the required register with the *D* register and ensure transport of the data.

The direct input and output activities of the memory (cycle stealing)

In the section describing the interfacing of the computer with the telecommunication system mention was made of the three input and output activites that are not initiated by a software instruction but by hardware. Known as cycle stealing, these three activities — dump, program load and housekeeping load — have priority over all other activities of the computer; the sequence in which they are mentioned here is also their relative order of priority.

The interruption of a cycle in operation is effected by the logic circuit in the control unit, in the case of dumping and program loading on the command of a 'request signal' delivered by *TMU* or *TCU* respectively, and in the case of housekeeping loading on the command of the 8 Hz clock pulse. The instruction being handled at the moment such a request appears is completed, and the *I/O* activity starts as soon as the execution-stage register *ESR* has returned to the rest state. If a dump request appears simultaneously with a load request the circuit takes the priority into account and therefore performs the dump first and then the loading operation. The loading of housekeeping data has the lowest priority.

In the next section we shall deal with memory selection and addressing, looking first at memory selection in the case of ordinary instructions.

[9] This compensation network, consisting of two thermistors and five resistors, generates a control voltage that compensates the measured drift of frequency with temperature. The network is optimized by first calculating the resistance values at which the sum of the squares of the deviations in the control voltage in relation to the ideal curve is a minimum. Measurements are then made on a network with these values to determine the compensation error resulting from the ohmic heating of the resistors and the non-ideal behaviour of the thermistors. Finally, these compensation measurements, which have an accuracy of 0.1 Hz in 8 MHz, are used to calculate the exact values of the resistors, which are then manufactured (the accuracy for these wirewound resistors is approximately 0.01 %).

Memory selection and addressing

The appropriate memory block is selected by the block selection circuit in the control unit *CU* (*fig. 8*). There are three block-selection registers in this circuit: *BSR/P* (3 bits), *BSR/DP* (3 bits) and *BSR/DA* (4 bits). By means of an electronic switch *S* the state of one of these three registers is presented to the block-decoding circuit in the memory, which determines in which of the seven blocks an input or output instruction is to be executed.

When the telecommand unit *TCU* receives a program-allocation command (*PAC*) from the ground station it loads into the register *BSR/P* three bits that indicate which block will serve as program memory during the next 12-hour interval. For the execution of an instruction requiring access to the program memory, the sequencer of the control unit sets the block-selection switch to state I and the contents of *BSR/P* can now be read in. Addressing within the block, i.e. within the program memory, takes place as described via the address register.

Reading and writing into the data memory requires the use of three bits to indicate in which of the other six memory blocks the operation is to be performed. This is done by means of the 4-bit register *BSR/DA*. This is a standard computer-output interface (one of the registers COI in fig. 6), located inside the onboard computer. This register is loaded with four bits by an output instruction; three bits are used for the memory-block address, while the fourth bit, called a 'flag bit', indicates whether a subsequent reading instruction is to refer to the program memory selected by *BSR/P* (flag bit 0) or to the part of the data memory selected by *BSR/DA* (flag bit 1). See also Table II. The flag bit thus determines whether the switch should be in state I or III. Addressing within the selected block again takes place via the address register.

For the dump action a dump allocation command (*DAC*) is transmitted to *TCU*, which then decodes three bits indicating the memory block from which data have to be dumped. These three bits are loaded into register *BSR/DP*. The switch is now set in position II and dumping is started upon receipt of a dump-request signal. Addressing within the block again takes place via the address register, but its state is now determined by a different operation register, the *dump-operation register*. Dumping of data from a particular memory block always begins at address 0; the dump operation register therefore begins in state 0. In the same way as for the ordinary operation register, the contents are incremented by 1 after each dump cycle, so that, once the dump request signal has been given, all 4096 words stored in the memory block are read out serially without interruption. This being done, a dump-ready signal is given to *TMU*, indicating that the dump action has been completed in this block.

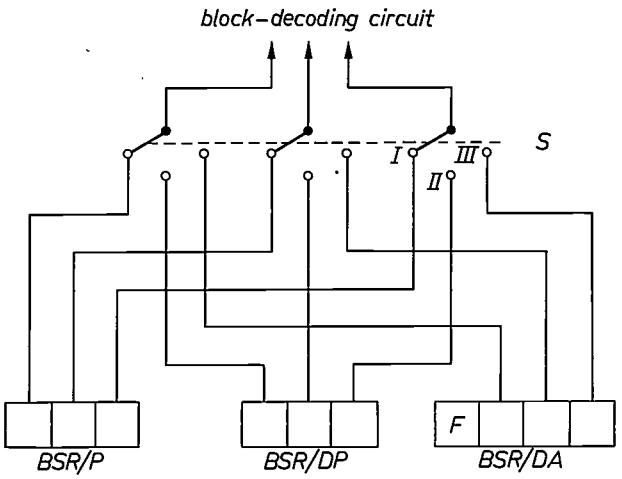


Fig. 8. Simplified diagram of the block-selection circuit of the control unit. The three registers *BSR/P*, *BSR/DP* and *BSR/DA* accept bits that indicate respectively which block is to serve as program memory, from which block data have to be 'dumped', and which block is to be selected for the data storage. The electronic switch *S*, operated by the control unit in fig. 6, passes the information to the block-decoding circuit in the memory distribution unit. The fourth bit of the register *BSR/DA*, the 'flag bit' *F*, indicates whether a read-out instruction is to be executed in the program memory or the data memory.

In loading a program or housekeeping data the block address is given by the contents of *BSR/P*, the data being loaded in the program memory. In program loading the addressing inside the block is done by the address register. As already described above, *TCU* then supplies words of 28 bits, consisting of 16 bits which are passed via the *M* register to the memory, and 12 bits which give the address. These address bits are read directly into the address register. In loading the housekeeping data one of the 64 permanent addresses in the

Table IV. Memory addressing in the onboard computer.

Computer activity	Block selection (3 bits)	Address selection (12 bits)
Software-controlled activities:		
fetch phase of an instruction	*BSR/P*	operation register
execute phase of an instruction relating to		
program memory	*BSR/P*	12 least-significant bits of instruction
data memory	*BSR/DA*	idem
Cycle-steal activities:		
dumping	*BSR/DP*	dump-operation register
program load	*BSR/P*	directly from *TCU*
housekeeping load	*BSR/P*	6 least-significant bits of divider chain in clock-pulse generator + 6 fixed bits

program memory, which are reserved for this purpose, must be selected. For this purpose the state of the six least-significant bits of the divider chain in the clock-pulse generator is transferred to the six last bits of the address register (these bits count 64 steps in 8 seconds), while the other bits of this register are filled with 000001 to define the location of the 64 addresses. A survey of the various modes of memory addressing for the various activities of the onboard computer is given in *Table IV*.

The memory

The memory comprises seven memory blocks (modules) and a central distribution unit. Each memory block contains a stack of 16 matrices of 4096 lithium-ferrite cores (diameter about 0.5 mm), organized by the '3D' method [10]. *Fig. 9* shows how the wires pass through the cores of a memory in this configuration. The memory distribution unit contains a pulse generator (timer), block-decoding circuits, address-decoding circuits, and a power-supply and distribution unit

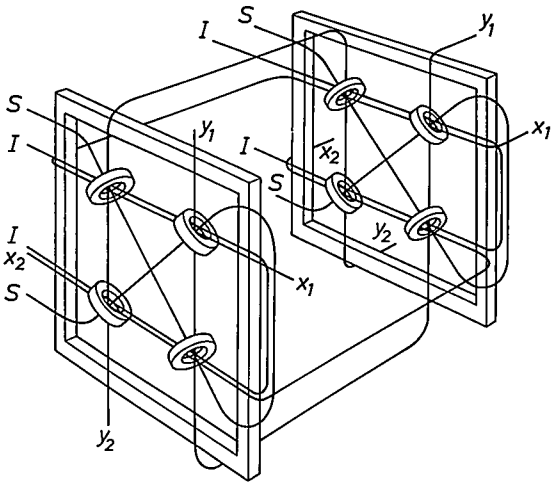


Fig. 9. Example showing the wiring in a '3D'-organized core memory. For simplicity, only two matrices with four cores are shown, i.e. a memory with a capacity of 4 words of 2 bits. There is an x and a y wire through each core for the read and write currents, a sense wire S on which the output pulse appears on read-out, and an inhibit wire I through which a current is passed on a write command if that particular core is to remain in state '0'.

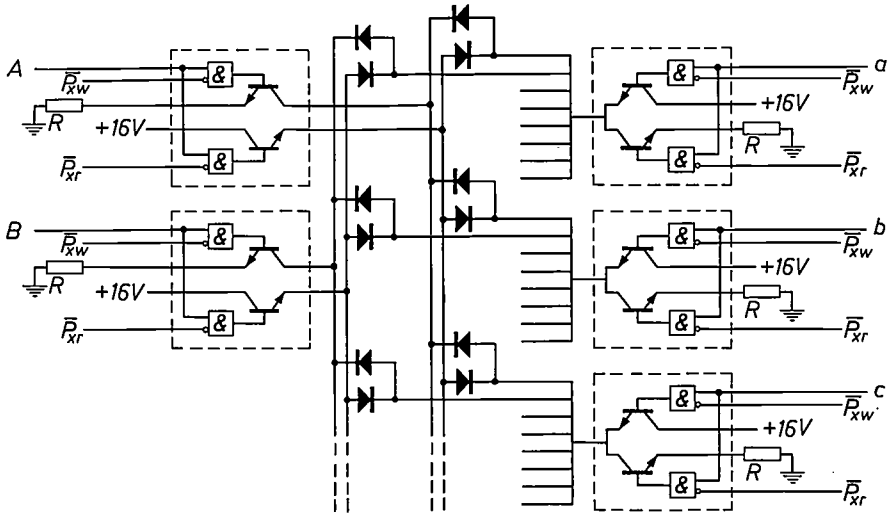


Fig. 10. Five of the 16 switches which, in two groups of eight, select the read and write currents through an x wire. A particular x wire is selected by activating one of the eight switches at both ends. If the write pulse P_{xw} appears (and thus the negation $\overline{P_{xw}}$ goes from 1 to 0), the two AND gates for which the signal 1 appears at the other input, open the switching transistors that drive the write current through the required wire. The input signals A, B, . . ., a, b, c, . . . are delivered by the address decoding circuit. Upon receipt of a read instruction the read pulse P_{xr} drives into conduction the two transistors that send a current in the other direction. The double selection switches (indicated in the figure by a dotted square) are hybrid integrated circuits specially designed for this application.

which produces a voltage of +16 V from the satellite supply voltage of 20 V for the read and write currents of the x and y wires in the blocks, and —5 V for the sense amplifiers. The +5 V supply of the satellite and the —5 V supply are switched on separately for each block and only when reading or writing action is necessary in the block. As described in the foregoing, this is determined by the block-decoding circuit.

To read and write into a block of $4096 = 64^2$ words it is necessary to select the x and y wires corresponding to the decoded address. This is done by means of 16 switches for the x wires and 16 for the y wires (see *fig. 10*). The required signals are supplied by the address-decoding circuit and the timer. Since each wire

[10] H. E. van Brück, Organization of ferrite core memories, Electronic Appl. Bull. **31**, 2-27, 1972.
H. J. Heijn and N. C. de Troye, A fast method of reading magnetic-core memories, Philips tech. Rev. **20**, 193-207, 1958/59.

is used for both reading and writing, which requires currents flowing in opposite directions, each 'switch' consists in fact of two switches, one for each direction; the currents are separated by diodes.

A complete read-write cycle takes place as follows (see *fig. 11*). On receipt of a start signal the timer generates the read pulses P_{xr} and P_{yr}, which close the switches in the read direction chosen by the address-decoding circuit. Next a signal P_{data} is sent to the M register of the arithmetic-logic unit, which then takes over the word read-out. The driving circuits of the inhibit wires (inhibit drivers) are now actuated, and the desired selector switches in the writing direction are then closed with the write pulses P_{xw} and P_{yw}.

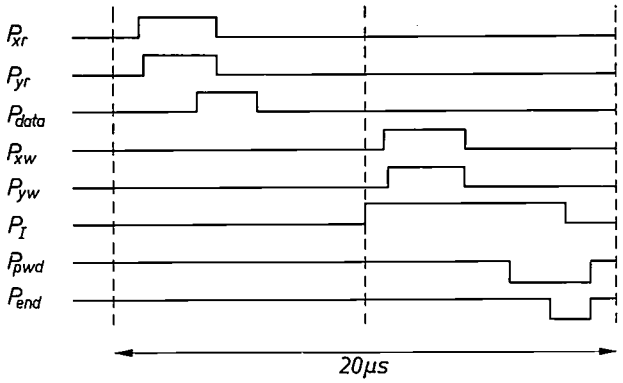


Fig. 11. Time diagram of the various signals that control a read-write cycle of the memory. The complete cycle takes 20 microseconds.

In each operation a number of cores are half-energized, which induces interfering voltages in the sense wires (S). To minimize these interfering signals there is a 120-nanosecond shift between the leading edges of the x and y read and write pulses. Furthermore, after a bit has been read in, a small demagnetizing pulse, called the 'post-write disturb' pulse, is applied via the inhibit wire to the cores that have just switched to the state 1. This slightly reduces the magnetization of these cores, and also helps to minimize the interfering signals induced by half-energization during subsequent memory cycles. These measures are of great importance here because in the matrix organization chosen each S wire passes through 4096 cores.

At the end of a read-write cycle a signal P_{end} is sent to the processor in the computer. The total memory cycle is completed in 20 microseconds.

Power supply and distribution

The onboard computer requires several different supply voltages. The central processor and the memory distribution unit only use a voltage of 5 V; the crystal oscillator uses this voltage and a voltage of 20 V; both voltages are provided by the satellite power-supply system. The x and y selector switches need 16 V and the sense amplifiers —5 V. These voltages are obtained from two power converters (d.c.-d.c. converters) one of which supplies the power for three memory blocks and the other for four, so that in the event of a defect in one of the converters a part of the memory will still function. The greatest proportion of the total power taken by the computer is dissipated in the memory. The power distribution is given in *Table V*, which also shows the importance of switching off a memory block when it is not in use. As can be seen, a total power of 6.6 W is required in a normal operation program.

Table V. Power consumption of computer during execution of a normal operating program.

Processor with crystal oscillators	2 W
Memory distribution unit	2.2 W
Memory distribution unit with one block switched on during memory activities	4.6 W
Memory with all seven blocks continuously switched on (not possible in practice)	26.7 W

Computer software

To enable it to perform the various tasks described at the beginning of this article, the onboard computer has an extensive package of software stored in the program memory. The programs come under four main headings:

1) Operation-execution monitor (*OEM*),
2) Attitude-control algorithms (*ACA*),
3) Experiment data handling (*EDH*),
4) Housekeeping data handling (*HDH*).

Together these programs are referred to as the operation execution program (*OEP*); they take up about three-quarters of the program memory, i.e. about 3000 words of 16 bits. Although it will not usually be necessary, it is possible to renew the entire operation-execution program from the ground, for instance in the event of an intermittent fault.

The *OEP* controls the execution of the activities that are defined in the satellite-operation program (*SOP*) for periods of 12 hours each (if necessary even for a period of 18 hours). *SOP* consists of a list of 16-bit words which indicate for each activity successively: a time (expressed in satellite-calendar time), a code word specifying the action which the satellite must carry out at that time, followed by the various parameters needed for this action. An *SOP* message of this kind for one period of 1 second (between two 1 Hz start pulses) is presented to the operation-execution program *OEP*. The execution of an *SOP* message can extend over a

Table VI. List and sequence of words which the satellite operation program *SOP* may contain in a message or instruction for a 1-second period

Mnemonic code	16-bit *SOP* word
TIWO	*Time word*, giving the satellite-calendar time at which the *OEP* must accept the code word COWO, and the subsequent parameters.
COWO	*Code word:* a collective word in which the following actions can be indicated: — the attitude submode to be used; — the scheme to be used for processing the measurement data; — the measurement command to be given to the observation systems; — the operating command to be given ('time-tagged' command); — the transition to the restricted mode, during which only a limited number of tasks are carried out (attitude control in the scanning mode and generating housekeeping data for the low-speed telemetry frame); — starting addresses for storage of measurement data. The code word also indicates whether the activity is to be carried out immediately or later, after the required attitude operating mode has been reached.
	Parameters
YCOG	y-coordinates and threshold values for the recognition of guide stars by the star sensor.
ZCOG	The same, for z-coordinates.
SHSE	Slew angle about the z-axis, defined as the angle between the positive x-axis and the local vertical, for turning the satellite into a new attitude.
STSY	Duration of the slew manoeuvre about the z-axis and the value of the offset angle about the y-axis.
SSSE	Setpoint for the y and z coordinates of the tracking star in the image field of the star sensor, so that the celestial object under observation enters the angle of aperture of the observation instruments such as the slit of the UV spectrophotometer.
SOFF	Magnitude of the offset angle in the z-direction and duration of the offset pointing, during the submode 'background measurement'.
SXOP	Operating code for the soft-X-radiation observation system (*SXX*).
UVCO	Measurement command to instrument measuring ultraviolet radiation (*UVX*).
SXCO	Measurement command to instrument measuring soft X-radiation (*SXX*).
HXCO	Measurement command to instrument measuring hard X-radiation (*HXX*).
STCO	Operating command
UVBA	Block number and initial address for storage of *UVX* measurement data.
SXBA	Block number and initial address for storage of *SXX* measurement data.
HXBA	Block number and initial address for storage of *HXX* measurement data.

longer period depending on the contents of the message. When the message has been executed, the computer returns to the 'soft-rest' state until a new *SOP* message is received. *SOP* is checked every second for the presence of a new message.

Table VII. Coding for two satellite activities with two 16-bit *SOP* words.

Activity	*SOP* words
Assignment of block numbers and initial addresses for storage of experimental results from the three observation systems	TIWO COWO UVBA SXBA HXBA
Slewing the satellite to a new attitude and performing fine pointing after recognition of guide stars. Issue of measurement command UVCO as soon as star-pointing submode is reached, and start of data processing as incated in COWO, after a time delay of a specified number of seconds.	TIWO COWO YCOG ZCOG SHSE SSSE STSY UVCO

Table VI gives a survey of the words that may form part of such a *SOP* message. The words are denoted by their mnemonic code (TIWO for time word, COWO for code word, etc.). The extremely compact coding of activities in these words enables a complete instruction package for a period of 12 hours, with a hundred activities to be executed, to be contained in about 1000 words. Because of this the *SOP* uses no more than one quarter of the program memory and can be stored along with the *OEP* in this memory. *Table VII* gives an example of the coding for two activities and a list of the *SOP* words, presented in their mnemonic code.

Fig. 12 surveys the tasks of the various parts of the operation-execution program *OEP* in executing an *SOP* message. The operation-execution monitor *OEM* is initiated by the 1-Hz start pulse. This program decodes the *SOP* message and governs the execution of the activities it contains by the other parts of the execution program. The *OEM* also monitors the state of the satellite and takes the necessary measures to deal with certain abnormal situations that can be recognized beforehand, for example if the satellite enters the Earth's shadow or in the event of loss of fine pointing.

After the attitude-control algorithms have brought the satellite into the correct attitude for the observations, the operation-execution monitor starts the measurements. The experiment data-handling program controls the read-out of the results from the data registers (*CII*) in the observation systems, where necessary carries out 'data reduction', and then stores the data in the data memory *DMY*. Read-out from a *CII* register takes place in two steps. First the required register is selected by means of an input instruction, and the information is then read out of this register and transferred to the *D* register. For a 16-bit word the whole process takes about 400 microseconds, which means that it is possible to repeat it many times per second.

Data reduction, which dispenses with redundant information, is used to save space in the data memory. The program contains four data-reduction schemes, relating respectively to the removal of the *eight* most significant bits, the removal of the *six* most significant bits, the removal of the *six* most and the *two* least significant bits, and the conversion of a number from the fixed-point notation into the floating-point notation, in which process a word of 16 bits is reduced to 8 bits (5 bits fixed-point part and 3 bits exponent of base number 2). The abbreviated words are then combined to form words of 16 bits.

The data from each observation system is stored in a prearranged part of the data memory. It is preferable to assign a separate memory block to each observation system. This makes it possible to make extra data dumps, enabling the contents of a memory block to be dumped in the interim to a ground station. This is important in the case of the X-ray observation systems when pulsar measurements with a time resolution of 1 or 4 milliseconds are carried out, since the data memory is then filled in about half an hour. For these extra data dumps the receiving equipment in some of the NASA ground stations will be used.

The housekeeping data-handling program *HDH* handles the data temporarily placed by *TMU* in the program memory. Some of this data is transferred by the software from the program memory to the data memory, from which it is dumped together with the experimental results during the next ground contact. Storage in the data memory takes place at fixed rates, which differ depending on the experimental data concerned and can be modified at each ground contact. In addition, in the event of certain abnormal situations, a series of relevant data can automatically be stored. The *HDH* program also generates at intervals of 1 second a word of 8 bits, which is transferred to *TMU* for inclusion in the low-speed telemetry frame. The contents of this word relate to the execution state of the program being handled at that particular second.

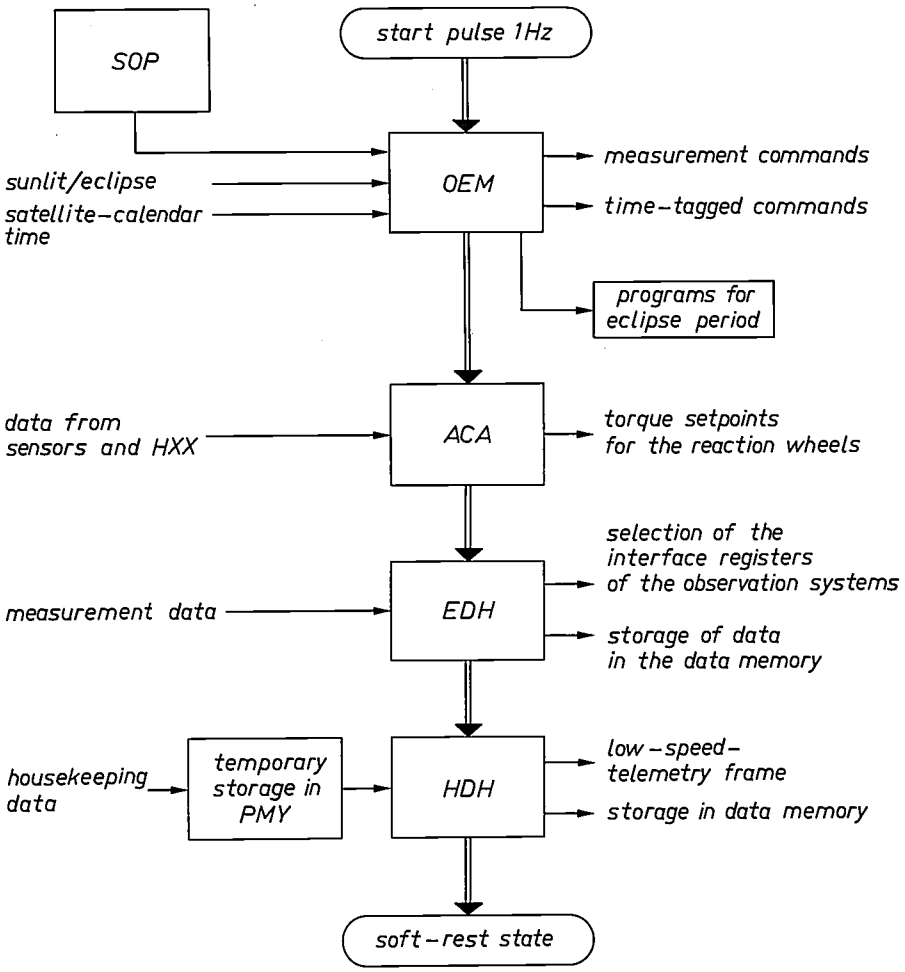


Fig. 12. Diagram of the computer software. On the appearance of a 1-Hz pulse the execution of a *SOP* (satellite operation program) message indicating a particular activity of the computer, initiates successively the operation execution monitor program (*OEM*), the attitude-control algorithm (*ACA*), the experiment data-handling program (*EDH*) and the housekeeping data-handling program (*HDH*). When the activity has been completed, the computer returns to the 'soft-rest' state. The double lines indicate a time sequence.

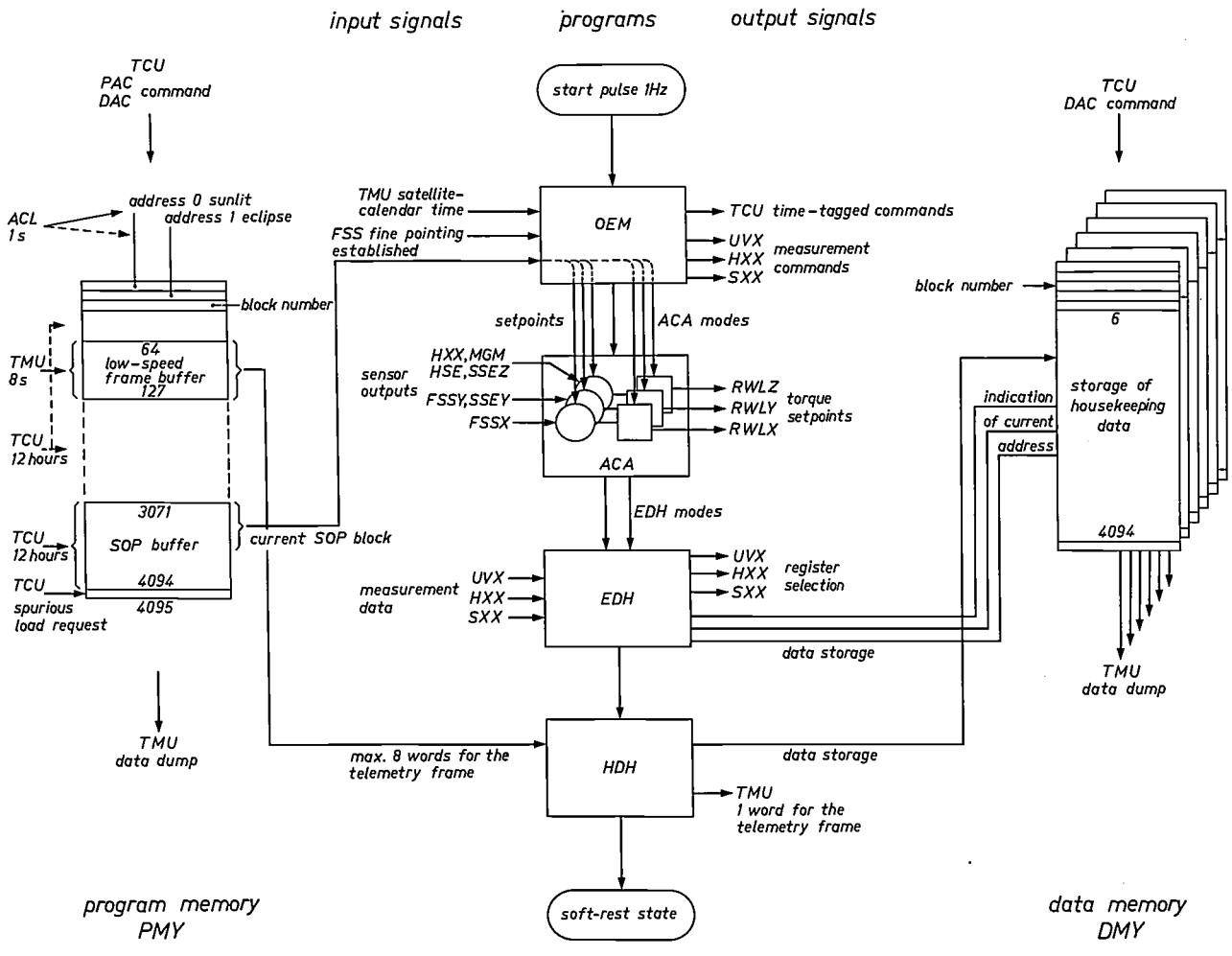


Fig. 13. Elaboration of fig. 11, showing the interrelations between the various programs, the program memory and the data memory. The abbreviations used have been explained elsewhere in this article

The program interfaces with the various memory blocks are shown in *fig. 13*, which is a more detailed version of fig. 12.

There are various software modes that correspond to various possible states of the satellite. They include the 'Sun-acquisition' mode, during which the computer only stores data on the operation of the satellite equipment, possibly related to events during this mode; the normal-orbit mode; an eclipse mode, which operates when the satellite no longer 'sees' the Sun; and a restricted mode that applies during a ground contact. In the restricted mode no experimental results from the astronomical equipment can be stored (*EDH* is then idle), nor is there any *SOP* decoding. The attitude control is carried out in accordance with the 'scanning' submode. The *HDH* program continues to supply data to *TMU* for the low-speed telemetry frame, which is now, however, not transmitted as such but incorporated in the dump signal. The program is put into the restricted mode by the last code word COWO. When the period of ground contact is nearly ended, this mode is terminated by a special memory-load message, which at the same time activates the newly loaded *SOP*.

Mechanical construction

The use of the computer in a satellite imposes very special requirements on the mechanical construction. Although compactness is also desirable in ordinary computers, the requirements imposed here as regards weight and insensitivity to vibrations and temperature fluctuations are out of the ordinary. To meet these requirements wide use was made of miniaturized components, and efforts were made right from the design stage to minimize the number of components. As far as possible existing components with a 'space qualification' were used; this qualification implies that the components meet a set of specifications for use in space such as ability to withstand vibrations in a frequency range from 20 to 2000 Hz with a maximum acceleration

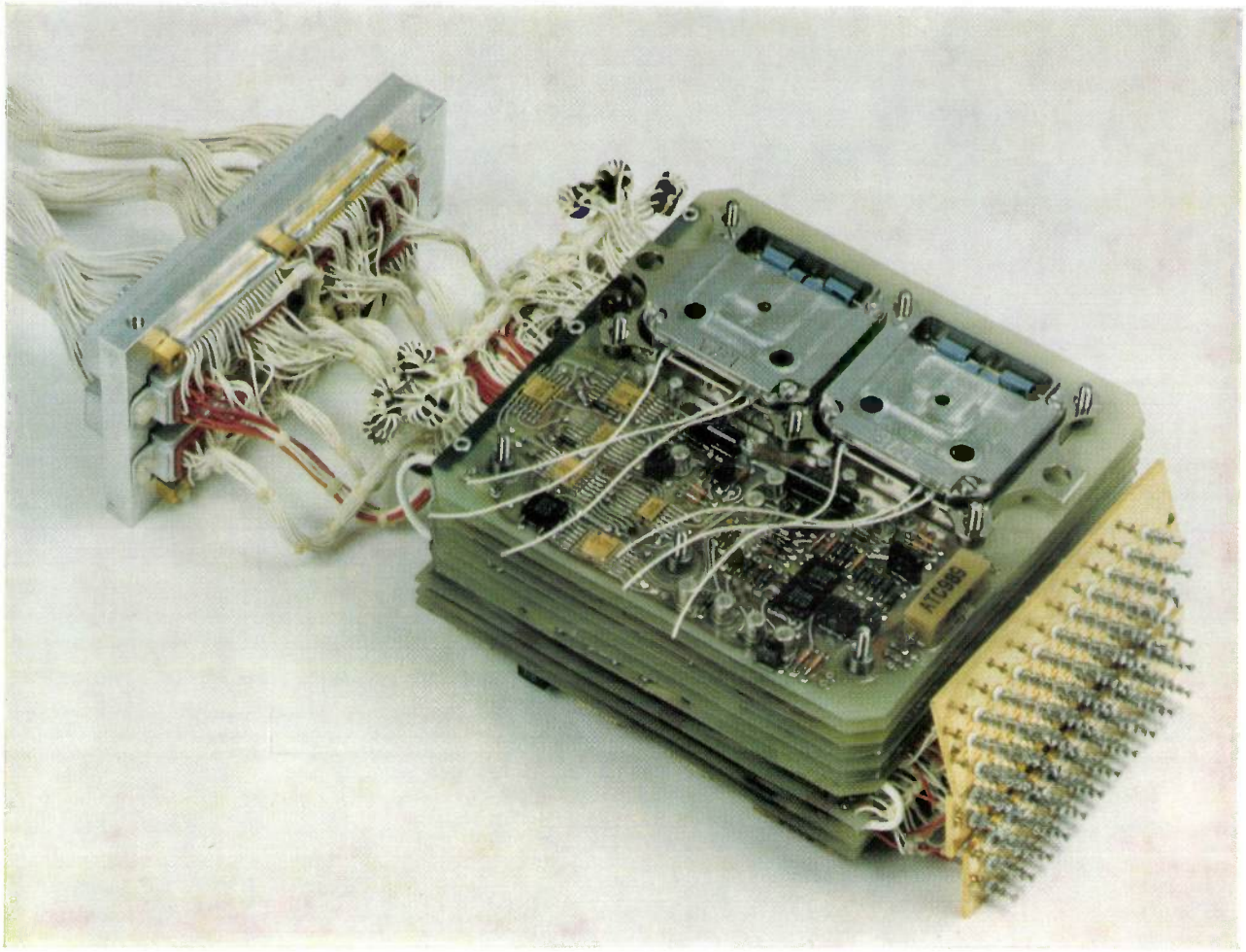


Fig. 14. The central module of the computer. The printed-circuit boards (dimensions 11.5×11.5 cm) contain the sub-units of the processor and the control unit of the memory. The outputs on the right serve for connections with the other modules of the computer; the connectors on the left make the connection with the other satellite systems.

of 70 *g*, and optimum operation in vacuum conditions at temperatures from −20 °C to +50 °C. The design of some of the units had to be modified in order to meet these specifications. The mechanical construction of the computer as a whole was subjected to very thorough tests under these conditions, and the operation of the computer was investigated in various combinations of unfavourable conditions (worst-case tests), for example at low temperature combined with a low supply voltage, both at the limit of tolerance.

The construction was based on well tried circuits and assembly techniques. The basic sub-units, mainly integrated circuits, are mounted on printed-circuit boards (dimensions 11.5×11.5 cm), which are joined together by titanium screws and nuts to form packaged units. The connections between the printed-circuit boards and between the modules of the computer are made by soldered joints, which the tests showed to be much more reliable than using strip-and-spring connectors. Adjacent boards are separated by aluminium spacers in which there are grooves filled with epoxy resin to give them some damping (this is another example of a modification developed during the tests). The various modules that make up the computer are mounted in eight units (dimensions 16×13×5 cm) milled from solid magnesium.

Fig. 14 shows the contents of the central module of the computer; it contains the processor with the crystal oscillators, together with the memory distribution unit. The processor contains about 160 integrated circuits, using low-power TTL logic. The memory distribution unit is made with multilayer printed-circuit boards, with six layers consisting of signal and supply-voltage layers separated by metal shielding layers.

The memory is also mounted on boards of 11.5×11.5 cm. *Fig. 15* shows the three boards carrying the core matrices and the diodes of the *x* and *y* wires (flat packs). The other circuits of a memory block are mounted on four multilayer boards fixed to the three boards mentioned above (see *fig. 16*). The various

boards are interconnected in such a way that the upper four can be unfolded for testing and maintenance purposes; the three boards of the memory stack form an inseparable unit. Hybrid thin-film circuits were specially developed for the x and y selection switches (*fig. 17*); each memory block contains 32 of these components.

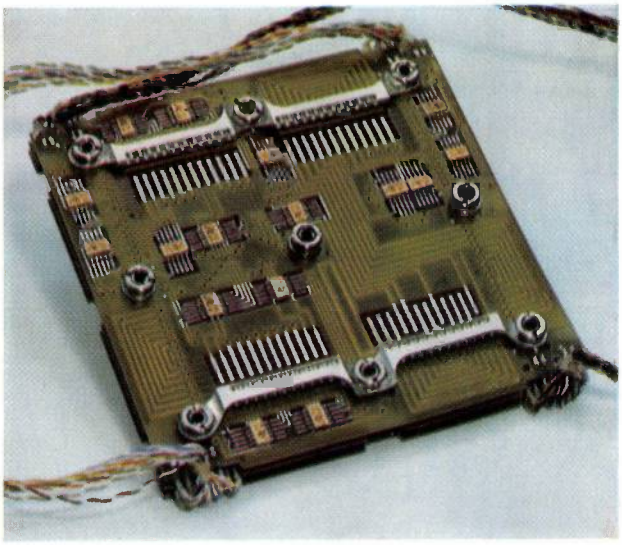


Fig. 15. Three printed-circuit boards with the core matrices of a memory block. The central board contains four matrices of 4096 cores on both sides (with 0.5 mm lithium-ferrite cores with an outside diameter of about 0.5 mm). The two other boards contain four matrices on one side (part of which may be seen in fig. 3) and interconnection patterns on the other side, while the upper panel also carries the diode arrays of the x and y wires. The three boards of the memory stack are inseparably connected by means of hollow screws and soldered strip connections along the edges. The connections to the other boards of the memory block are effected via wires in the four corners and via coaxial cables, whose sheaths are fixed in the four brackets visible on the upper board, with the wires soldered to the connections beside the brackets.

The connections between the eight modules of the computer are effected by means of a wiring pattern on one side of each module. Consequently, although soldered connections are used, the computer can nevertheless be opened up in the manner illustrated in fig. 3. Parts of the core matrices can be seen in the three memory modules shown on the left in this photograph) When the computer is completely assembled (see fig. 2. the wiring is covered by a surface shield. The modules are held together with two long screws through two holes at upper corners of the modules. Underneath they are fixed to two struts that also serve for mounting the computer in the satellite.

The design of the onboard computer was started in the middle of 1970, and by about the middle of 1971 the first provisional models of the processor and of the memory module were ready. These models were used for testing the parts of the software that had meanwhile been developed. Subsequently the first prototype was delivered, which contained only the processor and one memory module and was used with the single-axis model developed to simulate the attitude-control system [11]. The results obtained from the tests served as the basis for modifications to the computer, which were introduced in the next two prototypes, the development model *DM1* and the electrical model *EM*. Apart from being subjected to mechanical and electrical tests for space qualification, these models were used for elec-

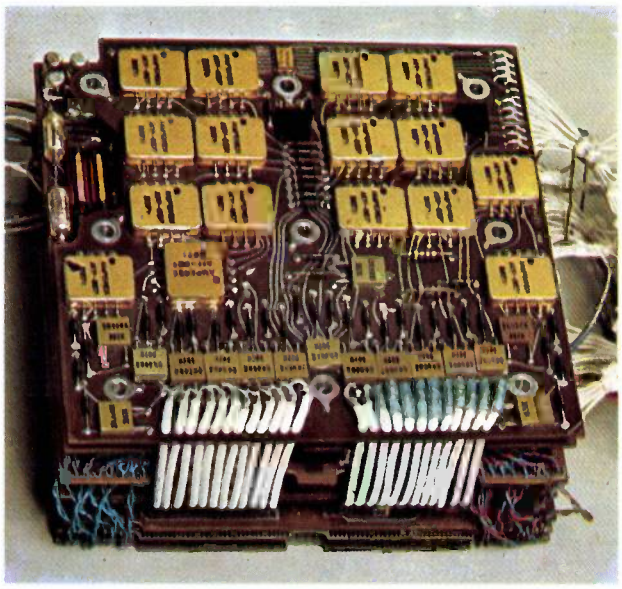


Fig. 16. A complete memory block. Mounted above the three boards with the matrices are four boards containing the inhibit drivers, the sense amplifiers and the x and y selection switches. The thick white wires are the coaxial cables to the matrices as indicated in fig. 15. The large gold-coloured blocks contain hybrid integrated circuits for the x and y wire selection.

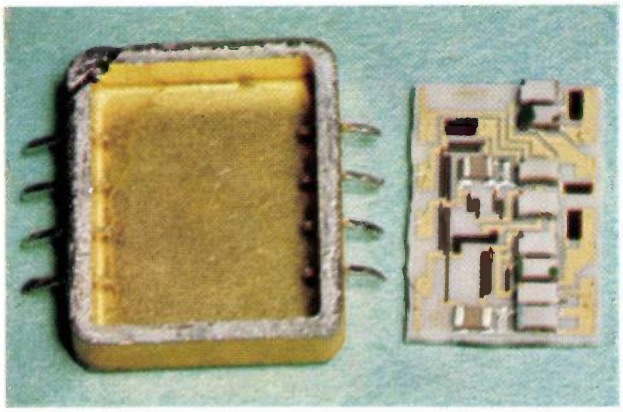


Fig. 17. A selection switch for the x or y current as in fig. 10, designed in the form of a hybrid integrated circuit. A metal can (see fig. 16) with outside dimensions of $12 \times 16 \times 3$ mm contains a ceramic substrate on which a pattern of conductors and resistors has been deposited by thin-film techniques. Discrete components such as transistors, diodes, capacitors and resistors are mounted on this substrate. The semiconductor components among them are mounted on small ceramic carriers soldered to the substrate upside down. The circuit contains a total of seven transistors. The selection switches were specially designed for use in the ANS onboard computer, and are made by Amperex (North American Philips).

[11] See the article of note [3], page 175.

trical tests together with prototypes of other components of the satellite. After a second development model *DM2* had been built to test the last modifications, a start was made in the autumn of 1972 on the construction of the two final models of the computer, one of which will be used in the satellite while the other will serve as a standby. The first was delivered in September 1973 and is at present being built into the satellite; when this is ready there will be a final series of tests before the launch.

Summary. The onboard computer of the Netherlands astronomical satellite (ANS) controls various systems in the satellite, including the observation instruments, during the 12-hour periods in which no contact is possible with the ground station, and stores the results of the observations in its memory. It is a digital computer with a core memory having a capacity of 28k words of 16 bits, divided into 7 blocks of 4096 words. One block works as the program memory and the other six as data memory. The memory has 0.5 mm lithium-ferrite cores in a '3D' organization; the memory cycle time is 20 μs. The computer uses the fixed-point notation; negative numbers are given in the two's-complement notation. The execution time for an addition is 128 μs. In addition to the A and M registers the arithmetic logic unit contains a third register which accepts the results of every addition in the other registers and which is also interfaced with the other systems of the satellite for the exchange of information between them. The control unit is responsible for controlling the various activities of the computer and for addressing the memory as well as for block selection within the memory.

There are three input and output activities of the computer that are not controlled by a program but by hardware. Known as 'cycle-steal activities', these are the loading of a new program during ground contact, 'dumping' the contents of the memory (with the experimental results) during ground contact, and loading the 'housekeeping data', which provide information on the state of the satellite systems. The housekeeping data is transmitted in the periods between ground contacts as tracking signals for locating the satellite.

The computer software consists of four main parts: an operation-execution monitor program, attitude-control algorithms, an experiment data-handling program and a housekeeping data handling program. At every ground contact a new satellite-operation program is transmitted in the form of about 1000 words of 16 bits, giving the instructions for the next 12-hour period.

Light materials such as magnesium and titanium were used for the construction of the computer to limit the weight to 8 kilograms. To keep power consumption below the specified upper limit of 8 W, low-power TTL logic is used, and the +5 V and −5 V supply voltages for a memory block are only switched on when the block is required for a read or write operation. Hybrid integrated circuits were specially developed for the x and y selection switches.

LOCMOS, a new technology for complementary MOS circuits

B. B. M. Brandt, W. Steinmaier and A. J. Strachan

Although the good characteristics of complementary MOS transistors have been known for some time, they have been very little used in LSI (large-scale integration) circuits because of the complicated processes and the low packing density. Now that the LOCOS technique is available CMOS transistors can be used to produce LSI circuits with high packing density and good electrical characteristics.

A type of MOS-transistor circuit arrangement that has now been known for several years is the 'complementary MOST circuit' (CMOS). The circuit has been given this name because it contains both *N*- and *P*-channel MOS transistors. Such circuits can have significant advantages over conventional MOS circuits [1]. The most important advantage is the low current level in logic circuits, giving a low dissipation. This enables static logic to be produced that would encounter considerable problems of heat dissipation if made entirely with *N*-channel or *P*-channel MOS transistors [2].

We shall explain this with the example of an inverter circuit. *Fig. 1a* shows such a circuit, made from CMOS transistors. With a positive supply voltage V_{dd}, if a positive voltage V_i is applied to the input (logic state '1'), then the *N*-channel transistor becomes conducting while the *P*-channel transistor does not conduct. The output voltage is then zero (logic state '0'). If the voltage is now removed from the input ('0'), the *P*-channel transistor becomes conducting and the *N*-channel transistor is switched off. The output voltage is now V_{dd} ('1'). The two MOS transistors therefore behave as switches, and the current in the circuit is determined by the very small leakage current of the switched-off MOS transistor. The current has a larger value only temporarily, during the switching, when there is some dissipation.

By way of comparison fig. 1*b* shows an inverter circuit made up from *P*-channel transistors [3]. The supply voltage here is $-V_{dd}$. If no voltage is applied to the input ('0') the lower MOS transistor, the 'switching transistor', does not conduct but the upper one, the 'load transistor' does. The output voltage is then equal to the difference between the supply voltage and the threshold voltage V_{th} of the load transistor: $-V_{dd} + V_{th}$ ('1'). When a negative input voltage is

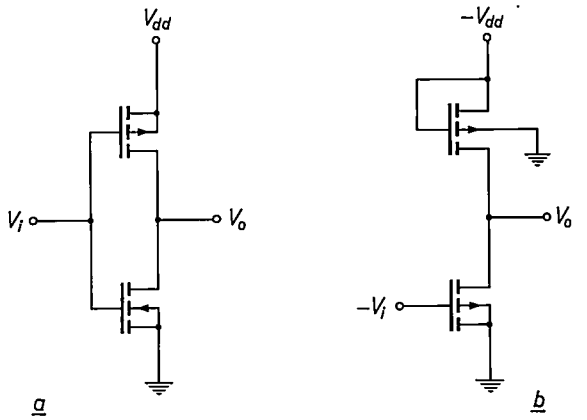


Fig. 1. *a*) An inverter circuit (schematic) consisting of an MOS transistor with a *P*-type channel *(above)* and another with an *N*-type channel *(below)*. The two gates are connected together, as are the two drains ('complementary' MOS transistors, CMOS for short). With a positive supply voltage V_{dd} and an input voltage V_i (logic state '1') the *N*-channel transistor conducts and the *P*-channel transistor does not. The output voltage V_o is now zero (logic state '0'). If $V_i = 0$ (state '0') the situation is reversed and the output voltage equals V_{dd} ('1'). The two transistors function as switches and in both states the only current in the circuit is the leakage current of one transistor.
b) Schematic circuit of an inverter circuit consisting of two *P*-channel MOS transistors, with the drain of one connected to the source of the other. With a negative supply voltage $-V_{dd}$ and a negative input voltage $-V_i$ ('1') both transistors conduct. The output voltage is then determined by the ratio of the channel resistances of the two transistors. If the channel resistance of the lower transistor, the 'switching transistor', is much smaller than that of the upper one, the 'load transistor', the output voltage is also very small ('0'). A current determined by the value of the channel resistances now flows in the circuit. If $V_i = 0$ (state '0'), the switching transistor does not conduct and the load transistor does. The output voltage is then equal to $-V_{dd} + V_{th}$ ('1'), where V_{th} is the threshold voltage of the load transistor. Only the leakage current of the switching transistor now flows in the circuit.

Ir B. B. M. Brandt, Dr W. Steinmaier and Dr A. J. Strachan are with the Philips Electronic Components and Materials Division (Elcoma), at Nijmegen.

[1] MOS transistors and circuits have been discussed in detail in Philips tech. Rev. **31**, No. 7/8/9, 1970 (the MOST issue).

[2] In static logic the information content of a logic circuit is held for an unlimited time, while in dynamic logic it is lost in a relatively short time. See also L. M. van der Steen, Digital integrated circuits with MOS transistors, Philips tech. Rev. **31**, 277-285, 1970. The differences between static and dynamic shift registers are discussed in this article.

[3] A more detailed treatment of this circuit is given in the article mentioned in note [2].

applied the switching transistor conducts. The output voltage is now determined by the ratio of the channel resistances of the lower and upper transistors. If this ratio has a small value, the output voltage is also small ('0'). But in this state there is a difference from the CMOS circuit: a current mainly determined by the channel resistance of the load transistor now flows through the transistors, so that there is dissipation. While this current can indeed be made small by making the resistance of the channel high, this can only be done at the expense of switching speed. The advantage of a CMOS circuit is that the channel resistance values can be small, and hence the switching speeds high.

Other advantages of CMOS circuits compared with ordinary MOS circuits are the immunity to fluctuations in the supply voltage or in the input voltage. The sensitivity to input-voltage fluctuations is low because the input voltage at which the circuit changes over from one logic state to the other is equal to about half the supply voltage, while the actual transition takes place over a very small range of input voltage. It is also easy to make a CMOS circuit compatible with other logic circuits such as DTL (diode-transistor logic) and TTL (transistor-transistor logic).

All these advantages would make CMOS transistors very suitable for use in integrated circuits, were it not for the fact that with the same tolerances the packing density is smaller than for ordinary MOS transistors. This means that the CMOS technique will give only a low yield when applied in large-scale integration (LSI circuits). Extra process steps are also required for a CMOS circuit, which has an adverse effect on the yield.

It has now been found that marked reduction in surface area can be obtained by using the LOCOS technique [4] developed at Philips Research Laboratories, combined with a special technique for applying *P*-type regions. This process is controlled in such a way that LSI circuits can be made.

In the LOCOS technique a silicon substrate is coated with a layer of silicon nitride, which is used as a mask in a later oxidation of the silicon when a silicon-dioxide layer is formed at the places where the nitride has been removed. Most of this 'LOCOS' oxide sinks into the silicon and gives good separation between regions of different doping. It takes up far less space than the conventional isolation diffusion. The dimensions of a circuit can be made even smaller since contact window and metallization masking do not have to be kept a certain minimum distance from the isolation diffusion, but can extend right up to the LOCOS oxide. In addition, narrow uninterrupted metallized tracks can now be applied, since there are no large steps on the surface due to oxide layers. Another advantage of the LOCOS technique is the low capacitance between the metallization and the silicon at the thick oxide layer; this enables fast switching speeds to be obtained.

We shall now describe the process used for making CMOS circuits by the LOCOS technique — we call the process the 'LOCMOS technique' — with the aid of *fig. 2*. The starting material is a wafer of *N*-type silicon whose surface has the ⟨100⟩ orientation. A surface with this orientation generally has very few surface states, and little charge appears in the oxide grown upon it; this gives a low and reproducible threshold voltage. The wafer is coated with a thin layer of silicon nitride, which is next removed at the places where the isolation oxide is to be formed, and the silicon is then oxidized until the oxide layer is 1.8 μm thick (fig. 2*a*). The next step is to remove the nitride at the places where the *P*-islands for the *N*-channel transistors have to appear; this is done by standard photo-etching techniques. After this *P*-type regions are produced at these places by a special technique (fig. 2*b*). In this technique the silicon is doped with boron in such a way that the boron concentration at the surface has the value necessary for good operation of the MOS transistor, while the maximum of the concentration profile is located about 1.5 μm beneath the surface. This approach prevents parasitic *N*-type channels from forming along the LOCOS oxide. With this method there is no need to use 'channel stoppers' — these are strongly doped regions included to counteract the formation of parasitic channels, and they take up a lot of space. After the *P*-diffusion the rest of the nitride is removed, and a thin oxide layer is formed thermally. A polycrystalline layer of silicon is then applied. Next the polycrystalline layer is doped with phosphorus to make it an *N*-type conductor, and a pattern is etched in it for the electrodes and a part of the interconnection pattern (fig. 2*c*); the doping is necessary to give a low series resistance of the conductors and hence a high switching speed. This treatment also gives a stable threshold voltage, since the phosphorous oxide produced binds sodium atoms and thus protects the silicon dioxide from atoms that are known to introduce mobile charge in the oxide. The next step in the process is to produce *P*-type sources and drains by boron diffusion at previously etched openings in the oxide layer (fig. 2*d*). The gates

[4] 'LOCOS' is an acronym from Local Oxidation of Silicon. A description is given in:
J. A. Appels, H. Kalter and E. Kooi, Some problems of MOS technology, Philips tech. Rev. **31**, 225-236, 1970;
J. A. Appels and M. M. Paffen, Local oxidation of silicon, new technological aspects, Philips Res. Repts. **26**, 157-165, 1971;
E. Kooi, J. G. van Lierop, W. H. C. G. Verkuijlen and R. de Werdt, LOCOS devices, Philips Res. Repts. **26**, 166-180, 1971.
See also Philips tech. Rev. **31**, 276, 1970.

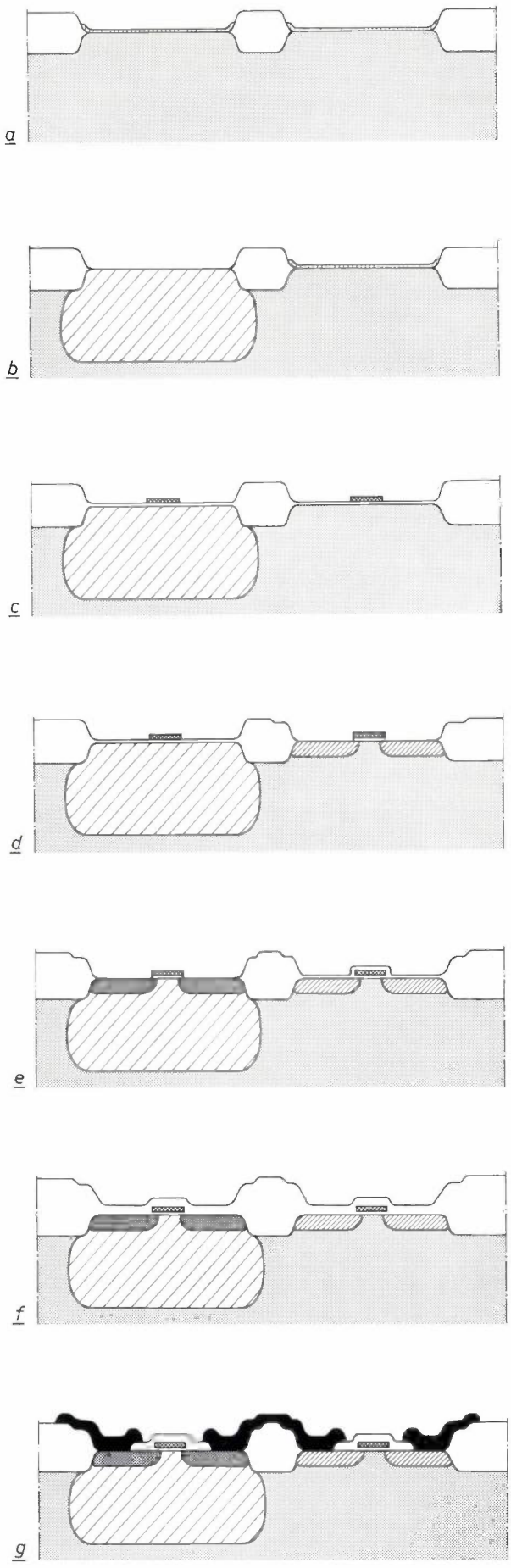


Fig. 2. The steps in the LOCMOS technique. *a*) The application of the LOCOS oxide. The *N*-type silicon is coated with silicon nitride in which openings are etched. The LOCOS oxide forms here as a result of an oxidation treatment. *b*) *P*-type regions for the *N*-channel transistors are made by diffusing boron through windows in the silicon-nitride layer. *c*) After removing the nitride, and forming a thin oxide layer on the silicon surface, a layer of polycrystalline silicon is applied. A pattern for the gates and their interconnections is etched in this layer. *d*) The sources and drains (P^+) for the *P*-channel transistors (*right*) are now formed by boron diffusion in the *N*-type regions, with the gates and the LOCOS oxide serving as a mask. *e*) The sources and drains (N^+) for the *N*-channel transistors (*right*) are now formed in a similar way by phosphorus diffusion in the *P*-type regions. *f*) An SiO_2 layer is next deposited pyrolitically, and openings are etched in the SiO_2 at the places where contact with the electrodes is required. *g*) An aluminium layer is then deposited by evaporation and the interconnection pattern for the circuit is etched in it.

oxide
nitride
polycrystalline Si
P-Si
P^+-Si
N-Si
N^+-Si
Al

and the LOCOS oxide serve as masks. Since these electrodes are small the stray capacitances are small, which also helps to give a high switching speed. After the boron diffusion a thin oxide is again formed on these regions. The *N*-type sources and drains are next produced in a similar treatment, with a phosphorus diffusion (fig. 2*e*). A silicon-dioxide layer is then deposited pyrolitically, and openings are etched in this to allow contact between the electrodes and the interconnection pattern (fig. 2*f*). Finally, a layer of aluminium is applied by vacuum evaporation and the interconnection pattern is formed in this by etching (fig. 2*g*).

The great saving in space obtained with the LOCMOS technique is demonstrated in *fig. 3*, which shows an inverter circuit made with this technique compared with the same circuit made with the conventional technique.

The LOCMOS process described here has been successfully applied in the manufacture of a number of integrated circuits. These include an inverter circuit, an 8-bit shift register and a static 256-bit random-access memory.

The inverter circuit has a delay time of 3 to 5 ns with a supply voltage of 5 V and an identical inverter circuit as load. Under similar conditions a conventional CMOS circuit has a delay time of at least 12 ns.

The 8-bit shift register has one series input and eight parallel outputs, which are all capable of driving one TTL input. This circuit operates up to a frequency of 10 MHz at a supply voltage of 5 V. The area occupied by this circuit is 2.5 mm^2; a conventional CMOS

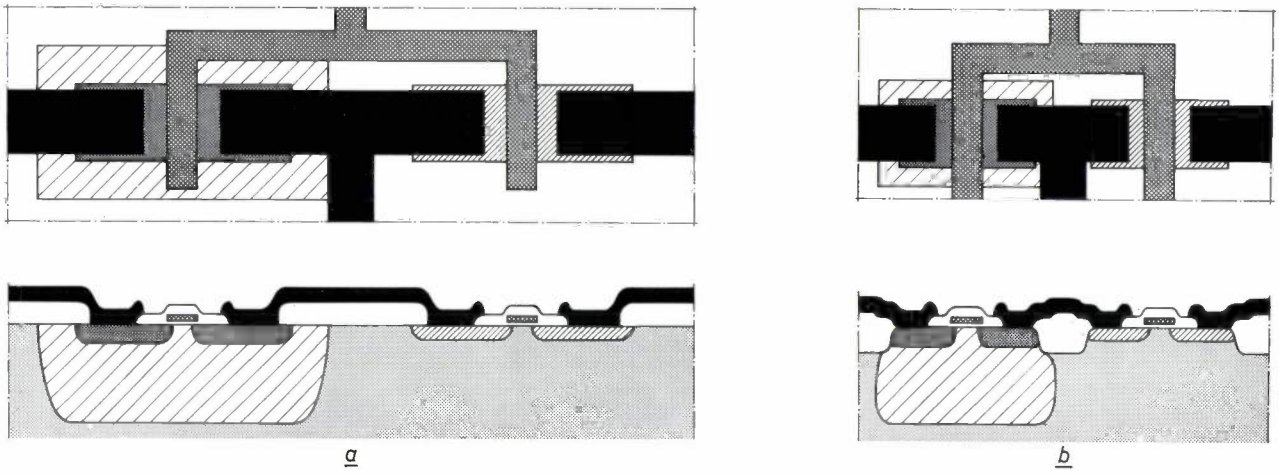


Fig. 3. Comparison of the dimensions of an inverter circuit made by the conventional process (*a*) and by the LOCMOS process (*b*). The structure is explained in fig. 2.

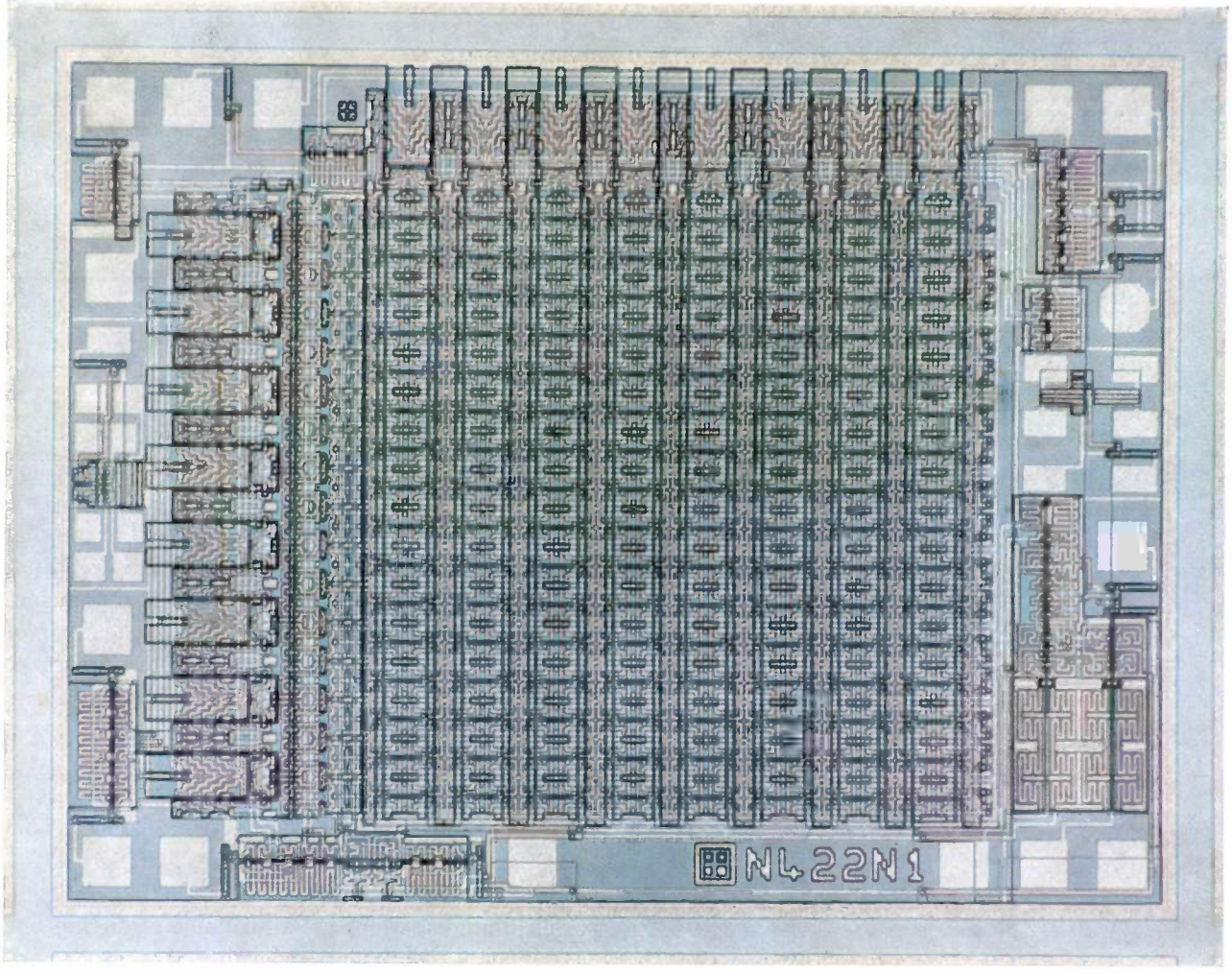


Fig. 4. A 256-bit static random-access memory made by the LOCMOS process. The circuit occupies an area of 5 mm².

circuit would occupy at least 5 mm^2.

The memory shown in *fig. 4*, occupies an area of 5 mm^2. The access time is 200 ns at a supply voltage of 5 V and is less than 100 ns if the supply voltage is increased to 10 V. The dissipation is extremely small. A similar circuit made without using the LOCOS technique would occupy more than 10 mm^2 and with a 5 V supply would have an access time of 600 ns.

These examples, particularly the last one, show that the use of the LOCOS technique has considerably increased the possibility of applying CMOS circuits in LSI.

Summary. Circuits with CMOS transistors have several good features, one of the most important being the low dissipation. It is however particularly difficult to apply them in LSI when conventional methods are used, partly because only low packing densities are obtainable. When the LOCOS technique is used, in which the silicon is locally oxidized by means of silicon-nitride masking, circuits can be made that have high packing densities and high switching speeds. Noteworthy features of this new LOCMOS technique are a special *P*-diffusion to produce a boron-concentration profile with a maximum *below* the silicon surface (to prevent parasitic *N*-channels from forming along the 'LOCOS oxide'), and the use of the LOCOS oxide and the interconnection pattern for the gates as masking for the formation of the sources and drains.

The article gives three examples of circuits made by the LOCMOS technique: an inverter circuit with a delay time of 3 to 5 ns, an 8-bit shift register occupying an area of 2.5 mm^2 and operating up to a frequency of 10 MHz, and a 256-bit static random-access memory with a surface area of 5 mm^2 and an access time of 100 to 200 ns.

Phosphors for the conversion of infrared radiation into visible light

J. L. Sommerdijk and A. Bril

Phosphors are generally used to convert radiation of relatively short wavelength (such as ultraviolet radiation) into visible light. Well known examples of this are the luminescent powders in fluorescent lamps. In these powders, Stokes's radiation law is satisfied; this states that the luminescent radiation has a longer wavelength than the exciting radiation. In recent years, however, phosphors have become known for which the situation is the opposite to that in the classical phosphors. These newer phosphors give a conversion from relatively long-wavelength (infrared) radiation to relatively short-wavelength (visible) radiation. The active ions in such substances gradually enter higher energy states by absorbing a number of infrared photons, and their return to the ground state is accompanied by the emission of visible light. By combining these phosphors with infrared-emitting diodes it is possible to make solid-state lamps that are suitable for many applications.

Introduction

Phosphors that can convert infrared radiation into visible light are of practical importance because they can be combined with gallium-arsenide diodes to form solid-state lamps; when a current flows in these diodes they give a very efficient emission of infrared radiation, which can excite the phosphors. Such a combination offers an alternative to the diodes that emit visible light when current flows, such as the GaP and SiC diodes [1].

To give an efficient conversion into visible light, both the infrared absorption of the phosphor and the efficiency of the conversion of the absorbed radiation into visible light must be high. A combination of ions of some of the rare-earth metals (RE ions) has been found capable of satisfying these requirements. This combination comprises trivalent ytterbium ions (Yb^{3+}) with trivalent ions of erbium, thulium or holmium (Er^{3+}, Tm^{3+} or Ho^{3+}), incorporated in a suitable crystal lattice. The Yb^{3+} ions can absorb the GaAs emission at a wavelength of about 1 μm, and are also able to transfer much of the absorbed energy to Er^{3+}, Tm^{3+} or Ho^{3+}. These ions can convert the energy obtained in this way to visible light. The ions Er^{3+} and Ho^{3+} can give either *green* or *red* light, depending on the host lattice in which they are located, while Tm^{3+} generally gives *blue* light.

In this article we shall first of all examine the mechanism of the conversions. Next we shall consider the factors that determine the colour of the emitted light and the efficiency. The combination of Yb^{3+} and Er^{3+} will receive most attention here, since it has been more fully investigated and gives the highest conversion efficiency. Finally we shall discuss the characteristics of the combination of a phosphor and a GaAs-diode and compare these with those of diodes that give a direct emission of visible light.

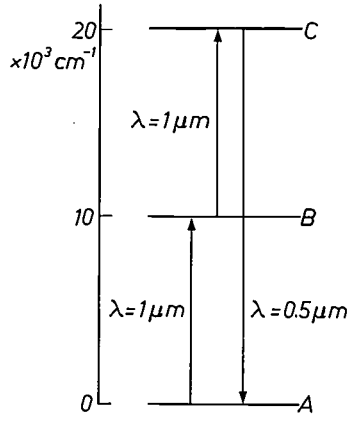


Fig. 1. Principle of the conversion from infrared to visible light. By successively absorbing two infrared quanta an ion can go from the ground state *A* via the excited state *B* to the state *C* provided the lifetime of the state *B* is long enough. The ion can then fall back from state *C* to the ground state accompanied by the emission of visible light. The wave number is shown on the left (this is the reciprocal of the wavelength, expressed in cm^{-1}, and is the energy scale normally used in spectroscopy).

Dr J. L. Sommerdijk and Dr A. Bril are with Philips Research Laboratories, Eindhoven.

Mechanism

The principle of the conversion of infrared radiation into visible light is illustrated schematically in *fig. 1*. Absorption of a quantum of infrared radiation (IR quantum) takes an ion from the ground state *A* to the excited state *B*. An essential condition for multi-step excitation is that the lifetime of this excited state should be sufficiently long for a second IR quantum to be taken up before the ion reverts to state *A*, enabling it to reach state *C*. Visible light will then be emitted as a result of the transition $C \rightarrow A$.

The ion must therefore possess energy levels at distances above the ground state that correspond to the energy of the infrared quantum (*B*) and to twice this energy (*C*), with the energy of the level *C* also corresponding to one quantum of energy in the visible spectrum. It is undesirable that there should be a large number of other levels between *A*, *B* and *C*, since most of the excitation energy would then be lost without any radiation. An energy-level diagram of this nature has only been found in the RE ions Er^{3+}, Tm^{3+} and Ho^{3+} [2].

The presence of these ions in a substance does not in itself make it a good phosphor. The long lifetime of the level *B* corresponds to a low probability for the transition $A \rightarrow B$, and hence to a low infrared absorption. These ions would therefore only be expected to give a small light output.

A significant increase in the light output can be obtained by adding the ion Yb^{3+}, mentioned in the introduction, to lattices activated with Er^{3+}, Tm^{3+} or Ho^{3+}. This effect was first observed by F. Auzel [3]. The Yb^{3+} ion gives an IR absorption about ten times stronger than that of Er^{3+}, Tm^{3+} and Ho^{3+}. In addition, the Yb^{3+} ions can transfer the absorbed energy efficiently to Er^{3+}, Tm^{3+} or Ho^{3+}, both when these are in the ground state *A* and when they are in the excited state *B*. It has also been found possible to incorporate more than ten times as much Yb^{3+} as Er^{3+}, Tm^{3+} or Ho^{3+} in a crystal lattice before there is any reduction in the light output because of interaction between neighbouring active centres. (We shall return to this later in the article.) The energy supply for the conversion of IR into visible light by Er^{3+}, Tm^{3+} or Ho^{3+} is thus completely controlled by the Yb^{3+} ions.

Fig. 2 shows the excitation routes for the conversion of IR into green and blue light. The conversion into *green* light is best for the combination Yb^{3+}-Er^{3+}. This requires two IR quanta, so that the intensity of the emitted green light increases as the square of the IR excitation density. Conversion into *blue* light is obtained with the combination Yb^{3+}-Tm^{3+}. To obtain one blue quantum, three IR quanta are necessary. The intensity of the emitted blue light therefore increases as the cube of the excitation density of the incident IR.

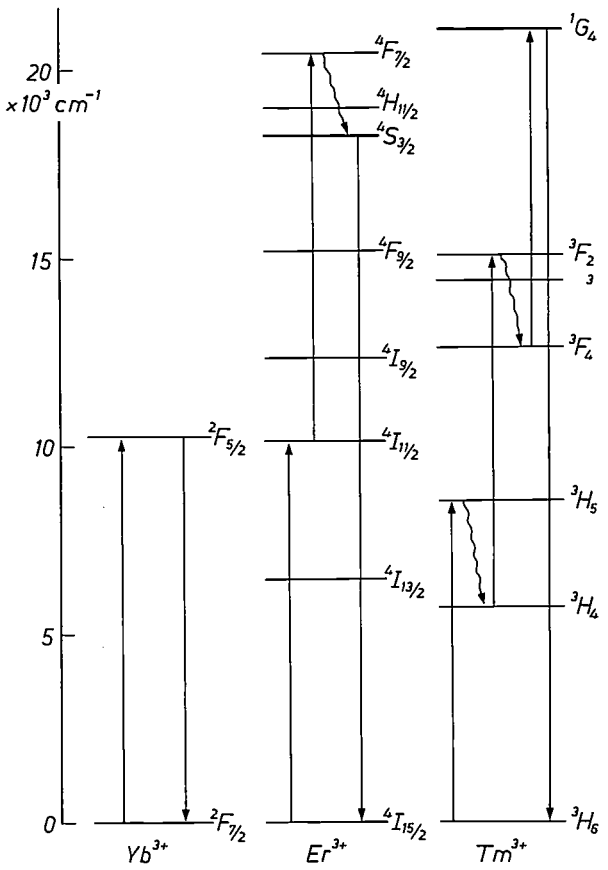


Fig. 2. Excitation routes of the ions Yb^{3+}, Er^{3+} and Tm^{3+} for the emission of green (Yb-Er) and blue light (Yb-Tm) on the absorption of infrared at a wavelength of about 1 μm. The infrared radiation excites the Yb^{3+} ions which then transfer the energy to the Er^{3+} or Tm^{3+} ions. Solid arrows correspond to transitions in which radiation is emitted, absorbed or transferred to neighbouring ions; 'wavy' arrows indicate non-radiative transitions, in which the energy is liberated as heat which is given up to the host lattice. The symbols by the energy levels give the quantum state appropriate to the electron cloud of the ions. Besides the azimuthal or orbital quantum number *L* of the total orbital angular momentum, the resultant spin quantum number *S* of the total spin angular momentum and the inner quantum number *J* of the total angular momentum are also of importance. These quantum numbers are summarized in the symbol $^{2S+1}L_J$ with the value of *L* given in code form in capital letters (S, P, D, F, G, H, I, etc. for $L = 0, 1, 2, 3, 4, 5, 6$, etc.).

The conversion into *red* light is again best with the combination Yb^{3+}-Er^{3+}; as we shall see later the host lattice determines whether green or red light is emitted. The conversion into red light can take place by various excitation routes (*fig. 3*). There are two routes in which two IR quanta are necessary for excitation (routes *a*

[1] More information about light-emitting diodes (LEDs) can be found in: R. N. Bhargava, Philips tech. Rev. 32, 261, 1971, or in: C. H. Gooch, Injection electroluminescent devices, Wiley, London 1973.

[2] A survey of the energy-level data for the RE ions is given in: G. H. Dieke, Spectra and energy levels of rare earth ions in crystals, Interscience, New York 1968.

[3] F. Auzel, C. R. Acad. Sci. Paris **262B**, 1016, 1966. This author has also published a survey article recently: Proc. IEEE **61**, 758, 1973 (No. 6).

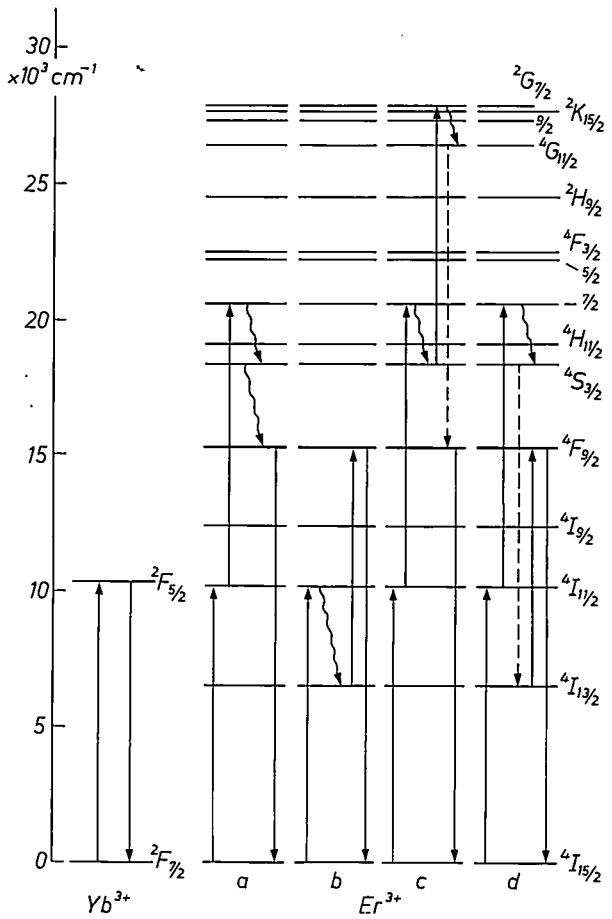


Fig. 3. Possible excitation routes for the emission of red light by the combination Yb^{3+}-Er^{3+}. The incident infrared radiation excites the Yb^{3+} ions, which then transfer the energy to the Er^{3+} ions. The dashed arrows indicate de-excitations in which energy is given back to Yb^{3+}.

and *b*), and two in which three quanta are necessary (*c* and *d*). We have found [4] that the most important route is not route *a*, the most obvious one, but route *b*.

For route *a* the $^4F_{9/2}$ level of Er^{3+} is populated via two non-radiative transitions from the $^4F_{7/2}$ level. The energy difference liberated in these transitions is given up as heat to the lattice. The relative unimportance of this route can be demonstrated from excitation spectra of the green and red emitted light with direct excitation in the wavelength range 350-500 nm. The green emission has a sharp peak corresponding to the excitation energy of the $^4F_{7/2}$ level (*fig. 4*). However, there is hardly any red emission for excitation to this level, which shows that the route given in fig. 3*a* is of no importance for the emission of red light.

For route *b* the absorption of the first IR quantum is followed first by a non-radiative transition to the $^4I_{13/2}$ level, and then by a second excitation step to the $^4F_{9/2}$ level. Population of the $^4I_{13/2}$ level is demonstrated by emission at 1.5 μm, due to the transition $^4I_{13/2} \rightarrow {}^4I_{15/2}$, observed on excitation with 1 μm radiation. Confirmation of the last step ($^4I_{13/2} \rightarrow {}^4F_{9/2}$) in route *b* is obtained from the increase in the red emission when the population density of the $^4I_{13/2}$ level is increased by simultaneous excitation at 1.5 μm.

In these routes two IR quanta are necessary to obtain one red quantum. Just as with the green emission, this leads to a square-law increase in the intensity of the emission with the IR excitation density. This square-law dependence is in fact found in all oxide lattices activated with Yb^{3+}-Er^{3+}. In fluoride lattices, on the other hand, IR excitation densities higher than 1 W/cm² give a greater increase in the red emission than corresponds to a square law. This is because the contribution of routes *c* and *d* to the total intensity is then no longer negligible. A red emission is obtained in this way after Er^{3+} has been excited three times by the absorption of IR quanta from Yb^{3+} (fig. 3). In route *c* Er^{3+} is excited from the $^4S_{3/2}$ level to the $^2G_{7/2}$ level, and then via a non-radiative transition to the $^4G_{11/2}$ level. Fig. 4 shows that the excitation to this level can give green as well as red emission. Red is emitted by preference, however, as can be seen from the emission spectrum obtained when the $^4G_{11/2}$ level is reached directly by excitation with long-wave ultraviolet (*fig. 5*). This selective de-excitation from the $^4G_{11/2}$ level is coupled with an energy transfer in which Er^{3+} gives back an IR quantum to Yb^{3+}, which is then excited to its $^2F_{5/2}$ level. This coupling becomes stronger with increasing Yb^{3+} concentration, resulting in an increase in the red emission.

While in route *c* the Er^{3+} ion from the $^4S_{3/2}$ level first takes up another quantum from Yb^{3+} and later gives a quantum back, exactly the opposite happens in route *d*. Here Er^{3+} first gives back a quantum to Yb^{3+}, so that it arrives at the $^4I_{13/2}$ state. From this level it is excited again to the $^4F_{9/2}$ state, exactly as in the last stage of route *b*. An indication of the de-excitation $^4S_{3/2} \rightarrow {}^4I_{13/2}$ is the reduction in the green emission on increasing the Yb^{3+} concentration, while at the same time the $^4I_{13/2}$ population increases, as indicated by the increase in the 1.5 μm emission at this higher Yb^{3+} concentration.

Finally, let us consider the decay times of the visible emission after removal of the IR excitation. In general these are appreciably longer than those relating to excitation with ultraviolet (UV) or bombardment with fast electrons. In these cases the decay times after removal of the excitation are mainly determined by the emptying of the levels that give rise to visible emission. After IR excitation, however, the decay times are determined by the emptying of the levels that give IR emission. This takes about 10 times as long as for the levels that give visible emission.

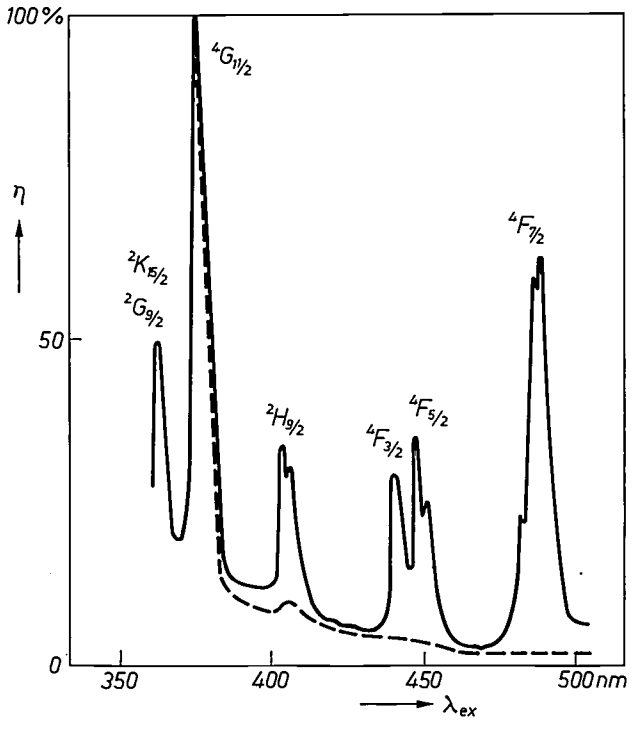


Fig. 4. Excitation spectrum for the green emission (solid curve) and the red emission (dashed curve) of α-$NaYF_4$: Yb^{3+},Er^{3+} for excitation by visible light. The relative outputs η of green and red fluorescent radiation are both plotted as a function of the wavelength λ_{ex} of the exciting radiation. The energy levels that give rise to the various emissions are given above the appropriate peaks.

Efficiency and colour of the emission

The efficiency of the conversion and the colour of the emission are found to be strongly dependent on the concentrations of Yb^{3+} and Er^{3+}, Tm^{3+} or Ho^{3+} and on the choice of host lattice in which these ions are incorporated.

It is important that these ions should be compatible with the host lattice. For this reason, compounds that contain the ions Y^{3+}, La^{3+}, Gd^{3+} or Lu^{3+} should be employed for the lattice. These ions can easily be replaced by Yb^{3+}, Er^{3+}, Tm^{3+} and Ho^{3+}, since the ionic radii do not differ greatly. Neither do these ions have any energy levels that can give emission or absorption in the infrared or in the visible range; they therefore take no part in the excitation processes we have described.

The compounds are chiefly oxides and fluorides that are prepared by solid-state reactions at high temperature. It has been found that fluorides are usually more suitable than oxides [5]. *Table I* gives a comparison of the emission intensities of the best fluoride phosphors with those of the best Yb^{3+}-Er^{3+}-activated oxides. It can be seen that the values for the fluorides are considerably higher than for the oxides. This is exactly the reverse of the situation with the UV-excited phosphors, in which oxides are usually more suitable than fluorides as host lattices [6].

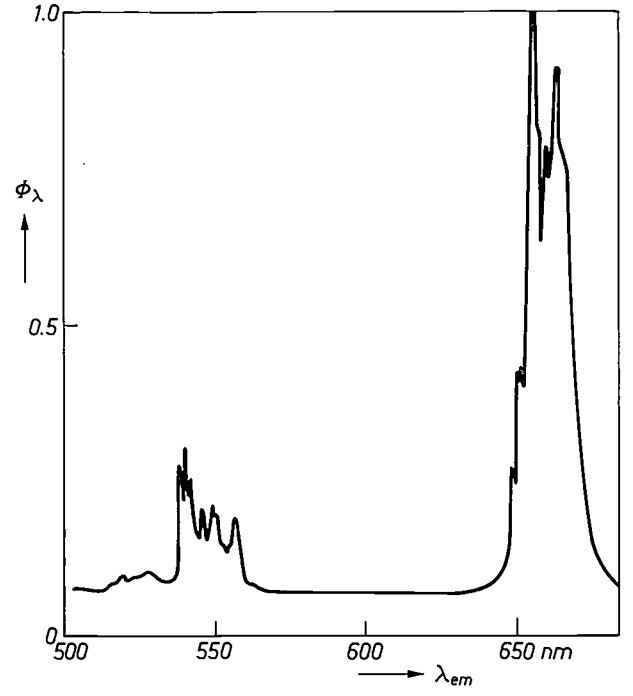


Fig. 5. Emission spectrum of α-$NaYF_4$: Yb^{3+},Er^{3+} for excitation with UV (wavelength 365 nm). The spectral radiant power Φ_λ is plotted in arbitrary units as a function of the wavelength λ_{em}.

Table I. The relative intensities of the emission of visible light by the combination Yb^{3+}-Er^{3+} on irradiation with infrared, for a number of fluorides and oxides as the host lattice. The particularly suitable host lattice α-$NaYF_4$ was discovered at Philips Research Laboratories by Dr G. Blasse and Miss A.D.M. de Pauw.

Host lattice	Intensity
α-$NaYF_4$	100
YF_3	60
$BaYF_5$	50
$NaLaF_4$	40
LaF_3	30
La_2MoO_6	15
$LaNbO_4$	10
$NaGdO_2$	5
La_2O_3	5
$NaYW_2O_8$	5

[4] J. L. Sommerdijk, J. Luminescence 4, 441, 1971, and J. L. Sommerdijk and A. Bril, Izv. Akad. Nauk SSSR, Ser. fiz. 37, 461, 1973 (No. 3), also published in: F. Williams (ed.), Luminescence of crystals, molecules, and solutions, Proc. Int. Conf., Leningrad 1972, p. 86.
[5] J. L. Sommerdijk, W. L. Wanmaker and J. G. Verriet, J. Luminescence 4, 404, 1971.
[6] G. Blasse and A. Bril, Philips tech. Rev. 31, 304, 1970.

This difference in behaviour can be explained from the difference in properties of the O^{2-} ion and the F^- ion [5]. The O^{2-} ion is much less stable than the F^- ion. The O^{2-} ion therefore transfers part of its charge much more readily to neighbouring cations, so that the bond between the ions is somewhat covalent in nature. This kind of transfer happens much less in fluorides, so that these are much more ionogenic in nature. The interaction between an activator ion and the host lattice is therefore much stronger in oxides than in fluorides. A strong interaction is advantageous in UV phosphors since interactions with the host lattice are essential here for the excitation of the activator ions [6]. In IR phosphors, however, a strong interaction with the host lattice is undesirable. It introduces a high probability of drop-out from excited states and thus leads to undesirably short lifetimes for the intermediate levels in the IR excitation process (e.g. level *B* in fig. 1). Oxides are therefore generally less suitable as host lattices for IR phosphors than fluorides.

Even within the two classes, fluorides and oxides, there is a marked variation [7] [8]. As an example of the variation with fluorides let us consider the results obtained for $NaYF_4$ activated with Yb^{3+} and Er^{3+}. (A common notation for this is $NaYF_4 : Yb^{3+},Er^{3+}$.) This compound can occur in two phases, the hexagonal α phase, stable at low temperatures, and the cubic β phase, stable at higher temperatures. The intensity and the colour of the emission with IR excitation are found to differ markedly for these two phases (*fig. 6*). The conversion efficiency is about 10 times lower for the β phase, in which the Er^{3+} ions are surrounded centrosymmetrically by the F^- ions. Because of this the emission transitions $^4S_{3/2} \rightarrow {}^4I_{15/2}$ (green) and $^5F_{9/2} \rightarrow {}^4I_{15/2}$ (red) have a low probability. Something similar has been found for UV excitation of Eu^{3+}, where the red emission, originating from the transition $^4D_0 \rightarrow {}^7F_2$, also has a low intensity in the lattices in which the Eu^{3+} ions are located at sites that are centres of symmetry [6].

There is however another reason for this relatively low intensity of the emission in the β phase. We have shown [7] that the interaction between the Er^{3+} ions and the surrounding F^- ions is much stronger in the β phase than in the α phase. As we saw earlier, a strong interaction does not favour a high conversion efficiency. However, when we compare the emission spectra of the two phases we see that the green emission is much more affected by the structure than the red. This is because the increased Er^{3+}-F^- interaction in the β phase also aids the non-radiative transition $^4I_{11/2} \rightarrow {}^4I_{13/2}$, which forms part of the main excitation route for the red emission (*b* in fig. 3). The decrease in the green emission consequently benefits the red emission to some extent. This increase in the red emission approximately compensates the decrease that arises for the two reasons mentioned in the last paragraph, so that in the end the intensity of the red emission does not differ so greatly for the two phases.

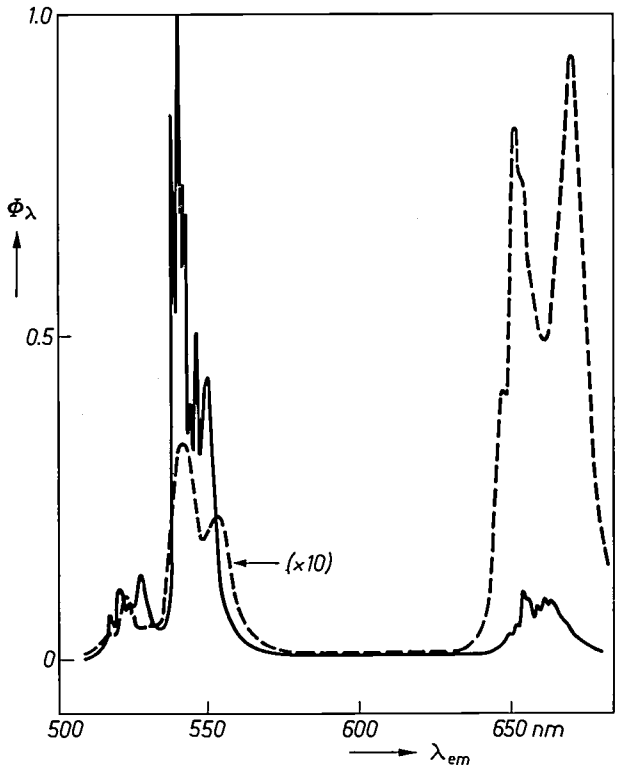


Fig. 6. Emission spectrum of the α phase (solid curve) and the β phase (dashed curve) of $NaYF_4 : Yb^{3+},Er^{3+}$ for IR excitation. The spectral radiant power Φ_λ is plotted in arbitrary units as a function of the wavelength λ_{em}. The spectrum of the β phase is shown on a scale ten times greater than that of the α phase.

In the oxides the intensity and colour of the emission vary strongly with the charge and the radius of the cations in the lattice [8]. The colour of the emission is mainly determined by the highest charge of the cations. This is shown in *Table II*. On increasing the highest cation charge from +3 to +6, the colour of the emis-

Table II. Effect of the highest cation charge in oxide host lattices on the intensity ratio of the green and red radiation emitted by the combination Yb^{3+}-Er^{3+} on irradiation with infrared. The second column gives examples of host lattices with the highest cation charge mentioned in the first column; the cation with this highest charge is underlined. The colour of the emitted light is given in the last column.

Highest cation charge	Example	Limits of the green/red ratio	Resulting colour
3	$\underline{Y}_2O_3$	0.0 - 0.1	red
4	$LiY\underline{Si}O_4$	0.2 - 0.4	orange, yellow
5	$LaN\underline{b}O_4$	0.5 - 5.0	green
6	$NaY\underline{W}_2O_8$		

sion changes from red via orange-yellow to green. This increase in the ratio of the intensities of the green and the red emission is a consequence of the decrease in the interaction between O^{2-} and Er^{3+} in the presence of cations of higher charge. It arises because on increasing the cation charge the O^{2-} ions become more strongly bound by these cations than by the Er^{+3} ions. The transition $^4I_{11/2} \rightarrow {}^4I_{13/2}$ therefore occurs less frequently and the red emission decreases in intensity with respect to the green. In some lattices containing ions with a positive charge of 6 this interaction is still strong in spite of the high charge, and these lattices therefore form an exception to the rule [8]. We have investigated the effect of the ionic radius in the system (La,Gd,Y) NbO_4 : Yb^{3+},Er^{3+}. Reducing the mean ionic radius by altering the relative proportions of La, Gd and Y resulted in a reduction of the total emission intensity, with the green emission falling away much more quickly than the red (*fig. 7*). This can also be attributed to the increasing O^{2-}-Er^{3+} interaction that arises if the O^{2-}-Er^{3+} spacing decreases as a result of the smaller radius of the ion to be replaced.

We have been able to confirm the various explanations for the changes in intensity and colour on changing the host lattice by measuring the decay time τ of the IR radiation after excitation with radiation at a wavelength of 1 μm [9]. It is found that in a number of cases the intensity of the green emission increases approximately as the square of τ (*fig. 8a*). As expected, we find a relatively low intensity in lattices in which the anions form a centro-symmetric environment for the Er^{3+} ions, such as $NaGdO_2$ and β-$NaYF_4$. A square-law relation between the intensity and τ, as found for the green, does not hold for the red (fig. 8*b*). For the same variation in the value of τ the intensity now does not vary by more than a factor of 10, as against a factor of 1000 for the green. This is because the various effects mentioned above tend to counteract one another here, so that the nature of the host lattice is of much less significance.

The intensity of the blue emission of Yb^{3+}-Tm^{3+} is found to be determined in very much the same way by the host lattice as the green emission of Yb^{3+}-Er^{3+} [10]. In *fig. 9* the intensity of the blue emission of Yb^{3+}-Tm^{3+} phosphors is plotted against the intensity of the green emission from the corresponding Yb^{3+}-Er^{3+} phosphors. The relation can be seen to be almost linear. Once again the highest intensity is obtained with the host lattice α-$NaYF_4$. The explanation for the lattice dependence is therefore completely analogous to the explanation for the green of Yb^{3+}-Er^{3+}. The combination Yb^{3+}-Ho^{3+} has been little investigated, since the intensities of the green and red emission are at least a factor of 10 smaller than for the combination

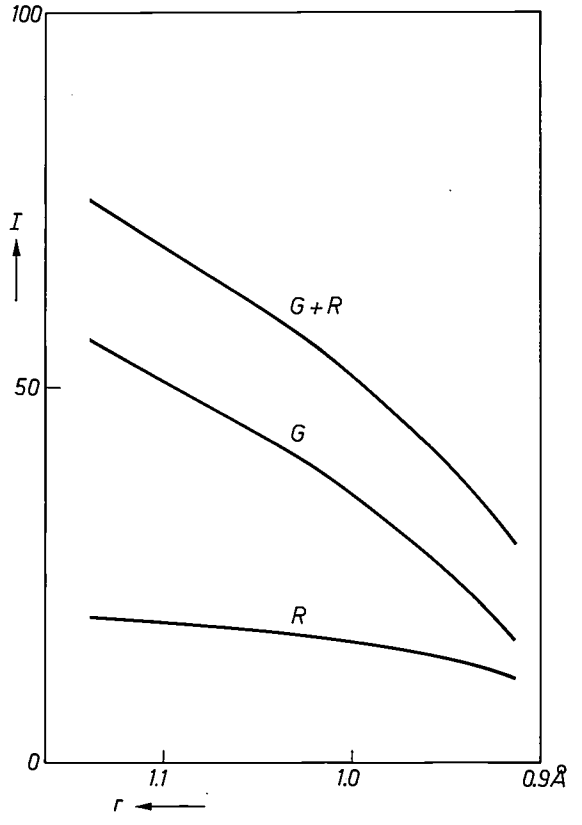


Fig. 7. The intensity I (in arbitrary units) of the visible emission of (La,Gd,Y)NbO_4 : Yb^{3+},Er^{3+} as a function of the mean ionic radius r. Curve R gives the red emission, curve G the green and curve $G + R$ gives the total emission for irradiation with infrared.

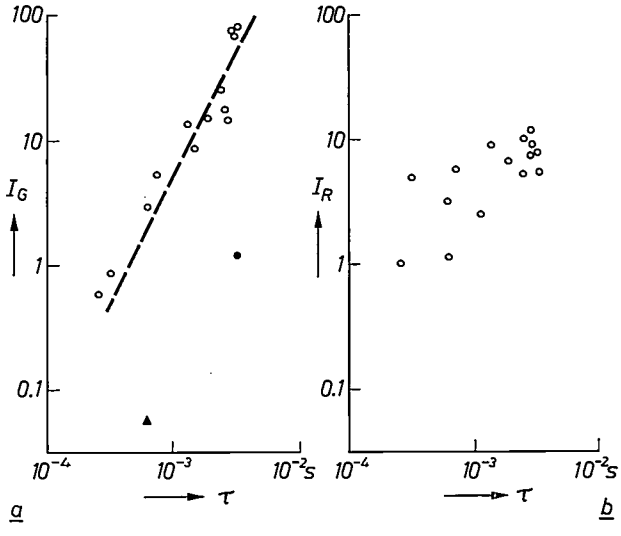


Fig. 8. *a*) The intensity I_G of the green emission of Yb^{3+}-Er^{3+} phosphors as a function of the decay time τ of the infrared emission. The host lattices mentioned in the text are indicated by different symbols; ● β-$NaYF_4$, ▲ $NaGdO_2$. *b*) As above, for the intensity I_R of the red emission.

[7] J. L. Sommerdijk, J. Luminescence 6, 61, 1973 (No. 1).
[8] J. L. Sommerdijk, W. L. Wanmaker and J. G. Verriet, J. Luminescence 5, 297, 1972.
[9] J. L. Sommerdijk, A. Bril, J. A. de Poorter and R. E. Breemer, Philips Res. Repts. 28, 475, 1973 (No. 5).

Yb^{3+}-Er^{3+}. It appears from the small amount of information available that the dependence on the host lattice is not very different here from that in Yb^{3+}-Er^{3+} and Yb^{3+}-Tm^{3+}.

Effect of the concentration

We shall now consider the effect of the RE concentrations in the various combinations. If Yb^{+3} ions are added in increasing concentrations to host lattices activated with Er^{3+}, then increasing numbers of IR quanta will be transferred to the Er^{3+} ions, causing an increase in the intensity of the green emission. At a particular Yb^{3+} concentration, usually somewhere between 20% and 40%, the intensity reaches a maximum; then as the concentration is increased further there is a gradual decrease in the intensity (*fig. 10, left*). This decrease in intensity is due to interactions in which, instead of luminescing, the Er^{3+} ions transfer a part of their energy to Yb^{3+} ions (fig. 10, *right*). These interactions increase with the number of Yb^{3+} ions in the neighbourhood of the Er^{3+} ions, and hence with the Yb^{3+} concentration. Correspondingly, if the Yb^{3+} concentration is increased, shorter decay times are therefore found for the green emission obtained on excitation with UV or fast electrons [9].

If the Er^{3+} concentration is gradually increased from zero, then the number of luminescing centres is increased at the same time. Here, however, a maximum in the green emission appears at a much lower concentration (about 2 to 4%), followed by a sharp decrease (*fig. 11, left*). This is caused by interactions between neighbouring Er^{3+} ions (fig. 11, *right*); both the strength and the number of these interactions increase very sharply as the Er^{3+} ions approach one another more closely. The considerable reduction in the decay times at higher Er^{3+} concentrations confirms this [9]. For the red emission of Yb^{3+}-Er^{3+} such intensity-reducing interactions are much weaker, with the result that at higher Yb^{3+} and Er^{3+} concentrations the red/green intensity ratio increases strongly.

For the combinations Yb^{3+}-Tm^{3+} and Yb^{3+}-Ho^{3+} the optimum Yb^{3+} concentration is again 20-40%, just as for Yb^{3+}-Er^{3+}. On the other hand the optimum Tm^{3+} and Ho^{3+} concentrations are only 0.1-0.3%. The concentrations are considerably lower than the optimum Er^{3+} concentration in Yb^{3+}-Er^{3+} because the interactions between neighbouring Tm^{3+} ions and neighbouring Ho^{3+} ions are much stronger than in the case of Er^{3+}.

Finally, we should mention the purity of the starting materials in the preparation of the phosphors discussed above. A compound of any particular RE ion is usually to some extent contaminated by other RE ions. These 'foreign' RE ions in the phosphor will absorb a part of

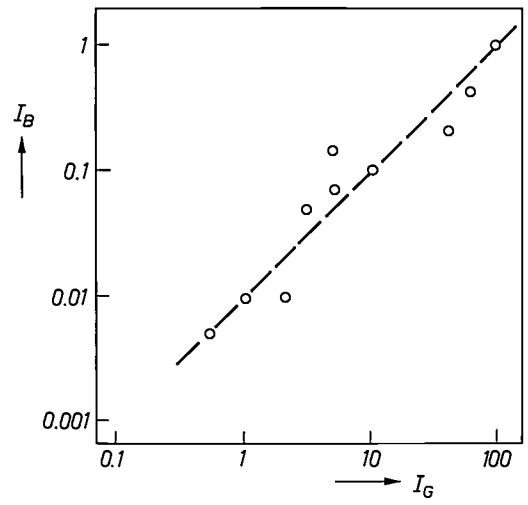


Fig. 9. The intensity I_B of the blue emission of Yb^{3+}-Tm^{3+} phosphors as a function of the intensity I_G of the green emission of Yb^{3+}-Er^{3+} phosphors with the same host lattice. In both cases the infrared excitation density is about 100 mW/cm². The intensity of the green emission here is about 100 times as great as that of the blue.

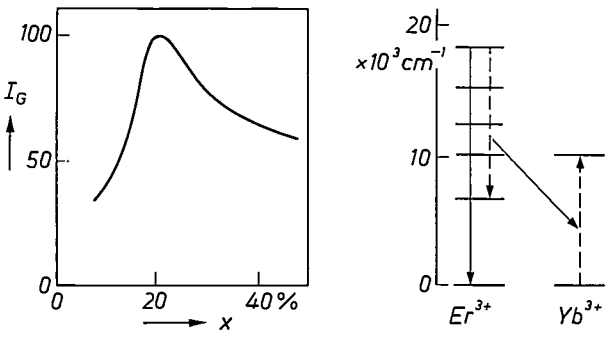


Fig. 10. *Left:* The intensity I_G of the green emission of α-$NaYF_4$: Yb^{3+},Er^{3+} for irradiation with infrared, as a function of the Yb^{3+} concentration x. The Er^{3+} concentration is 3%. *Right:* The energy-level diagrams of Er^{3+} and Yb^{3+}; a thin arrow indicates which energy transfer can take place from an Er^{3+} ion to a neighbouring Yb^{3+} ion. On increasing the Yb^{3+} concentration this transfer eventually dominates the beneficial effect of the Yb^{3+} ions on the infrared absorption. This is the cause of the fall in the green emission at higher Yb^{3+} concentrations.

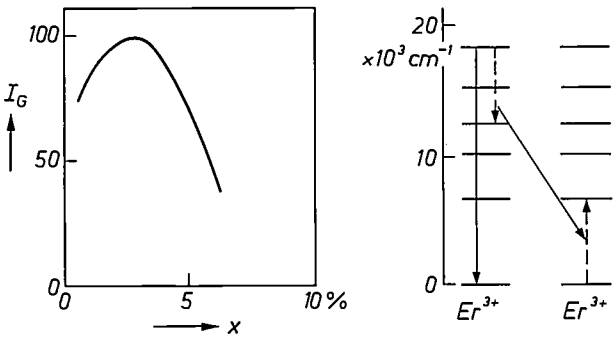


Fig. 11. *Left:* The intensity I_G of the green emission of α-$NaYF_4$: Yb^{3+},Er^{3+} for irradiation with infrared, as a function of the Er^{3+} concentration x. The Yb^{3+} concentration is 20%. *Right:* The energy transfer of an excited Er^{3+} ion to a neighbouring non-excited Er^{3+} ion. This transfer causes a sharp fall in the green emission if the Er^{3+} concentration becomes too high.

the irradiated IR energy. Since the foreign ions possess all the undesirable infrared levels mentioned at the beginning of the article, the energy absorbed can disappear via non-radiative transitions.

The concentration of the undesirable RE ions should therefore usually be no higher than about 10^{-3}%. This means that the RE oxides used as starting materials (Y_2O_3, La_2O_3, Gd_2O_3 and Lu_2O_3) must be at least 99.999% pure. The purity of Yb_2O_3 can be a little lower, 99.99%, while for Er_2O_3, Tm_2O_3 and Ho_2O_3 99.9% is adequate. For lower purities there is a noticeable reduction in the luminescence. For example, the presence of no more than about 0.01% of Dy^{3+} reduces the green emission of Yb^{3+}-Er^{3+} by a factor of ten.

Combination of IR phosphors with a GaAs diode

A light source can be made from an IR phosphor and an IR-emitting diode by coating the diode with a layer of the phosphor [11]. The IR phosphor is applied as a powder with a small quantity of binder (e.g. glycerine or an epoxy resin). The most suitable size of particle for the powder is found to be about 10 μm. The optimum thickness of the phosphor layer is about 200 μm; with a thinner layer too little of the IR from the diode is absorbed, and with a thicker layer the phosphor itself absorbs too much of the light produced within it. The most suitable GaAs diodes for our purposes have been doped with Si [12] and emit IR with a maximum intensity at about 940 nm. The IR excitation spectrum of the phosphor usually has a maximum at 970 nm. The wavelength region in which these two spectra overlap is not large (*fig. 12*). The maximum of the IR emission from a GaAs diode can be displaced to a longer wavelength by increasing the Si content of the GaAs [13]. Calculations show however that the improvement obtained from this displacement is cancelled out by a broadening of the IR spectrum and a decrease in the efficiency of the diode.

To see how high an efficiency can be attained, e.g. for the green, with the GaAs-phosphor combination, it will be useful to consider successively the four steps in the complete process. These are the IR emission from GaAs, the IR absorption of the phosphor, the excitation of the phosphor and the green emission from the phosphor.

The IR emission of a planar GaAs diode has an external efficiency of 5 to 10% [11] [12]. This efficiency is limited by strong internal reflection at the surface because of the high refraction index (3.5) of GaAs. This internal reflection can be counteracted by using a convex surface; the external efficiency can then be raised to 30% [12]. In combination with an IR phosphor, however, such an arrangement gives little improvement. While there is indeed more IR radiation from the GaAs, the excitation density of this radiation becomes much lower because of the increase in surface area. Since the light output is proportional to the square of the excitation density, as we saw at the beginning of this article, this means that there is hardly any improvement. We therefore have to work with a maximum diode efficiency of about 10%.

The IR absorption of the phosphor is also about 10% at the most under favourable circumstances. The rest of the radiation goes right through the phosphor or is reflected back into the GaAs to be absorbed there.

The excitation of the phosphor to the $^4S_{3/2}$ level increases with the square of the quantity of IR radiation absorbed. In principle, provided that this quantity is large enough, very nearly all the IR absorbed can be used for the excitation of the Er^{3+} ions to the $^4S_{3/2}$ level (fig. 2).

Finally, we should say something about the green emission of the phosphor. Once the Er^{3+} ions are in the $^4S_{3/2}$ state, a part of the stored energy is converted into green radiation by the transition $^4S_{3/2} \rightarrow {}^4I_{15/2}$ (fig. 2). A significant amount is lost, however, as a result of the interactions described earlier between Er^{3+} and Yb^{3+} ions or between the Er^{3+} ions themselves (fig. 10, fig. 11). The amount lost can be found by looking at the decay time of the green emission after excitation with UV or by fast electrons [9]. At very low Er^{3+} con-

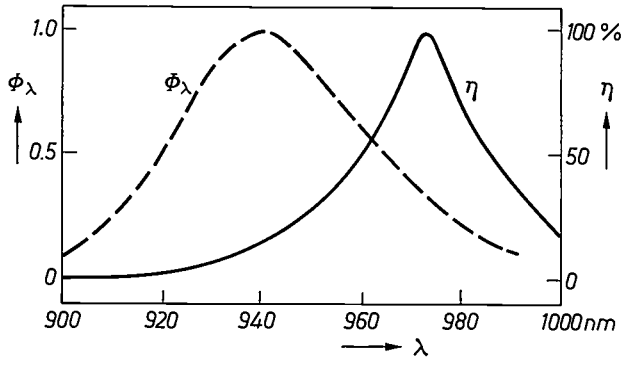


Fig. 12. The IR emission spectrum for silicon-doped GaAs diodes (dashed line) and the excitation spectrum for the green emission of α-$NaYF_4$: Yb^{3+},Er^{3+} (solid line). The spectral radiant power Φ_λ of the diode and the relative output η of the phosphor are plotted in arbitrary units as a function of the wavelength λ.

[10] J. L. Sommerdijk, J. Luminescence 8, 126, 1973 (No. 2).
[11] S. V. Galginaitis and G. E. Fenner, in: Gallium arsenide, Proc. 2nd Int. Symp., Dallas 1968, p. 131.
[12] These diodes are made by epitaxial deposition of GaAs from a solution in molten Ga; both the *P* and *N* layers are obtained by doping with Si ('amphoteric doping'). The deposition temperature determines whether the Si is incorporated at a Ga or at an As site, i.e. whether the material is *N*-type or *P*-type. Details are given in: H. Rupprecht, J. M. Woodall, K. Konnerth and D. G. Pettit, Appl. Phys. Letters 9, 221, 1966, and also in the book by Gooch of note [1], p. 83.
[13] I. Ladany, J. appl. Phys. 42, 654, 1971.

centrations (about 0.1%) and in the absence of Yb^{3+} the decay time τ in say α-$NaYF_4$ has a value of about 1 ms. This corresponds to a transition probability ($1/\tau$) of about 1000 s^{-1} for the green emission. The decay time for a sample of the optimum composition (about 20% Yb^{3+}, 3% Er^{3+}) is only 0.1 ms. This reduction by a factor of ten is caused by the competitive non-radiative processes described earlier. The efficiency of the green emission from a sample of the optimum composition is therefore no greater than 10%.

Taking all four steps into account, we arrive at a maximum value of about 0.1% for the overall efficiency. This is also approximately equal to the efficiency obtained with GaP light-emitting diodes that convert electrical energy directly into green light [1] [3]. To achieve an efficiency of 0.1% with a diode-phosphor combination, a relatively high current density is required (at least 100 A/cm^2). With GaP diodes the same efficiency can be attained at a tenth of the current density. However, at this current density the efficiency of the GaAs-phosphor combination is no more than 0.01%. One advantage of the combination is that it emits an attractive colour of green; the maximum of the very narrow emission spectrum (see fig. 6) coincides approximately with the maximum of the sensitivity curve of the eye. We therefore maintain that the GaAs diode in combination with a phosphor as source of green light is still to some extent competitive with diodes that emit light directly.

For the red, however, the situation is less favourable. Here again the efficiency of a GaAs diode combined with a phosphor is 0.1% at most, while with GaP diodes values at least ten times higher can be attained [1]. The maximum efficiency for the blue is estimated to be 10 to 100 times lower than for green and red. This efficiency is very low (10^{-2}-10^{-3}%) because three IR quanta are now required in the excitation process to obtain the desired emission (fig. 2). Even so, the efficiency for blue is still higher than that given by an SiC diode [3].

Summary. The mechanism of the conversion of infrared radiation into visible light by phosphors depends upon the presence of ions that have three or more energy levels spaced at distances that can be bridged by an infrared quantum. The lifetime of the intermediate levels must be large enough to allow a reasonable occupation of the highest level to be reached eventually. The rare-earth ions Er^{3+}, Tm^{3+} and Ho^{3+} are found to satisfy these and a number of subsidiary requirements. To obtain sufficient infrared absorption from the phosphors, Yb^{3+} ions must also be incorporated, which transfer the energy absorbed directly to the luminescing ions. The colour of the emitted light is determined by the particular ions and to some extent by the host lattice in which the ions are incorporated. The host lattice also has a considerable effect on the efficiency of the phosphors. The overall efficiency of the phosphors in combination with infrared-emitting GaAs diodes is 0.1% at most for the emission of green and red light. For green this combination can to some extent compete with direct light-emitting diodes, but not for red. The efficiency of the diode-phosphor combination is only 10^{-2}-10^{-3}% for blue, but this is still better than for a diode emitting blue directly.

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 2/3

Previous issues of our journal have included articles about the Netherlands astronomical satellite (ANS), due to be launched in August 1974, many of whose systems have been produced within the Philips group of companies. These systems include the attitude-control system, the telemetry equipment and the power-supply units. In these preceding articles little was said about the scientific objectives to be pursued with the satellite or about the instruments on board. In the two articles presented here, which have recently been published in the Nederlands Tijdschrift voor Natuurkunde (the Dutch Journal of Physics), these subjects are dealt with at some length by the two Dutch groups of astronomers who have designed the instruments and will direct the research to be made with them. We are much indebted to the Chairman of the Editorial Board of the NTvN, Prof. H. L. Hagedoorn, and to the six authors for their permission to reproduce these articles and for their kind cooperation.

The Groningen ultraviolet experiment with the Netherlands astronomical satellite (ANS)

J. W. G. Aalders, R. J. van Duinen and P. R. Wesselius

Introduction

Observations throughout the centuries through the 'optical window' of the atmosphere have given us a picture of a wide variety of astronomical objects and phenomena. From near-neighbours in our solar system and from remote galaxies the Earth receives photons of visible radiation that provide information about the physical conditions at the place of origin and about the spatial interrelationships of the sources.

In spite of the vast quantity of observations and the impressive body of knowledge built up from them, the picture obtained by optical astronomy is still very incomplete. Just how incomplete our knowledge is of the structure of our Milky Way system, for example, was demonstrated in particular by radio astronomy, developed after the Second World War. Opening this new window on the universe around us led to discoveries that were completely unexpected.

Dr R. J. van Duinen leads the Photometry working group at Groningen University. He is the Local Project Manager for the ANS project. Drs J. W. G. Aalders and Dr P. R. Wesselius are members of this working group.

It seems as if new insight is gained whenever additional observations become available from previously inaccessible parts of the electromagnetic spectrum. Thus it was found that systems exist that produce X-radiation [1], a discovery that led to speculation on the existence of 'black holes'. In another wavelength region, the infrared, unexpected radiation is found from cool dust aggregations surrounding clusters of hot stars.

Astronomical observations were extended from the visible region to the ultraviolet when it became possible to put space probes into a ballistic orbit at heights sufficient to enable observations to be made — though only very briefly — in this wavelength region, which had previously been inaccessible because of atmospheric absorption. From the visible observations it had long been known that very high temperatures occur in certain stars. Through observations in the ultraviolet it was possible to observe for the first time the most intense part of the spectrum of such hot stars

[1] See for example the article by A. C. Brinkman, J. Heise and C. de Jager in this issue, page 43.

($T > 10\,000$ K), which formed an extremely valuable addition to the existing knowledge. The space-probe experiments have meanwhile been followed by experiments with telescopes in man-made satellites. Several such satellites with instruments on board have already been launched. The Netherlands satellite ANS is one of the next to be put into orbit [2].

Two kinds of observations, classified by the technique employed, may be distinguished among the space experiments carried out so far. The first category relates to spectroscopic observations on bright stars, performed with high spectral resolution. Experiments of this type are particularly important in the ultraviolet, since in many atoms and ions the transitions to and from the ground state are accompanied by emission and absorption of radiation in the ultraviolet. Insight obtained into the structure of the star and detailed knowledge of the physical conditions during the formation of absorption lines are very important for testing theoretical models.

The second category relates to observations with a relatively low spectral resolution. Since this category of experiments may involve faint stars, observations can be made of so many objects that the information obtained can also be used as a means of improving the criteria used for classifying stars, and also for studying the distribution in space of certain types of object.

The ultraviolet observations to be carried out with ANS belong to the second category. They differ from similar experiments performed previously in that the satellite and our instruments which it carries allow accurate photometric measurements to be made on very remote stars that are relatively close to one another. To accomplish this the sensitivity has been increased by about 100 times with respect to that of previous experiments and the field of view used is so small that the radiation which the spectrophotometer receives from the celestial background is at least an order of magnitude weaker than that of the observed object. The field of view chosen is 2.5×2.5 minutes of arc, which implies that the pointing accuracy of ANS must be ± 1 minute of arc, a stiff requirement for such a small satellite. To achieve this with the size of telescope available, it is obvious that a great deal of attention had to be paid to the transmission of the measuring system and to the photon-detection efficiency. The small field of view of our measuring instrument and the very high pointing accuracy of the Netherlands satellite allow the stars of a cluster to be investigated one by one, which was not possible with the two previous ultraviolet experiments.

We shall now go somewhat more deeply into the scientific programme, and we shall continue with a description of the instrument constructed and the considerations underlying its design. *Table I* presents the principal data for our ANS experiment and the two previous UV experiments.

Table I. Comparative data for the ANS UV experiment and the two similar satellite experiments made previously.

Satellite	Telescope diameter	Wavelength range	Field of view	Sensitivity (m_v)	Launch
OAO-A2	20 cm	1200-4000 Å	10′	8th magnitude	Dec. 1968
ESRO/TDIA	27 cm	1350-2750 Å	12′ × 17′	7th magnitude	March 1972
ANS	22 cm	1500-3300 Å	2.5′×2.5′	11th magnitude	Aug. 1974

Observation programmes

The primary aim of the Groningen UV experiment is to measure the absolute intensities of a large number of weak objects at five wavelengths. Measurements will be made on open clusters and stellar associations, on hot subdwarfs, on short-period variable stars, on interstellar and circumstellar matter, on extragalactic nebulae and on various other celestial objects. Most programmes link up closely with observations made from the Earth in the visual region and with results obtained from the two satellite experiments mentioned in Table I.

Many programmes relate to very hot objects, for the following reason. To a first-order approximation the atmosphere of a star has one temperature and thus behaves as a perfect radiator (black body). According to Planck's radiation law, objects that are hotter than about 10 000 K radiate most energy in the ultraviolet, which cannot penetrate the Earth's atmosphere. Obviously, however, these hot objects can best be studied from observations of their ultraviolet radiation, which is why we need instruments carried by a satellite. We shall discuss a few programmes at somewhat greater length below.

The main programme is directed towards a better classification of these hot stars. Their classification in the visible region is based on the use of two obser-

vational parameters, one a measure of the absolute luminosity $L(\lambda)$ in a specified wavelength region and the other the effective temperature T_e of a star. Diagrams in which quantities such as absolute visual magnitude are plotted as a function of the spectral type (*fig. 1*), known as Hertzsprung-Russell diagrams, reveal a distinct concentration of points in particular regions. The diagram shows principal groups of stars, referred to as the main sequence, the giants, etc. The classification for stars with temperatures higher than 10 000 K is very rough, because such stars look much the same in the visible region. Any appreciable refinement of the existing rough classification must therefore depend on measurements in the ultraviolet. The five bands of the ANS ultraviolet instrument have been chosen at the wavelengths considered most suitable for such a classification. In our five-colour photometry we define two parameters B_1 and B_2, which are a measure of L(UV) and T_e. Both B_1 and B_2 must of course be calibrated against L(UV) and T_e.

The calibration is made in the following way. The distance to one cluster of stars, the Hyades, has been determined very accurately by geometrical methods. The intensity of a number of stars in this cluster is determined by observation, and this, combined with the distance, gives the energy flux in the ultraviolet leaving the star (L(UV)). The Hyades cluster only occupies a small region of the main sequence. Since we do not in the first instance know how the main sequence continues towards higher temperatures in the ultraviolet, we look for another cluster of stars whose main sequence partially overlaps that of the Hyades, and we fit the two partial main sequences together as well as possible. Some five star clusters are needed to complete the main sequence. One of these clusters, the Pleiades, is shown in *fig. 2*. The values of T_e with those of L(UV) then have to be found by comparison with model calculations.

This calibration is now applied to other hot stars. Those stars that have the same parameters B_1 and B_2 as a calibration star are assigned the same values of L(UV) and T_e. Whether this is possible without risk of confusion should appear from a comparison with models; the wavelength bands have of course been chosen in a way that assumes that it is. With this improved classification we hope to be able to determine the distance to stellar clusters containing a few hot stars more accurately than in the past. We shall try, for example, to improve on the distance to the h and χ Persei clusters, which contain four Cepheids. The exact distance to the Cepheids is very important, because the astronomical scale of distances is based on these variable stars.

Another interesting region in the Hertzsprung-Russell diagram in fig. 1 is the empty band in this figure at $T_e > 10\,000$ K and M_{pg} between −5 and 5. A study of the weak blue objects, especially in directions per-

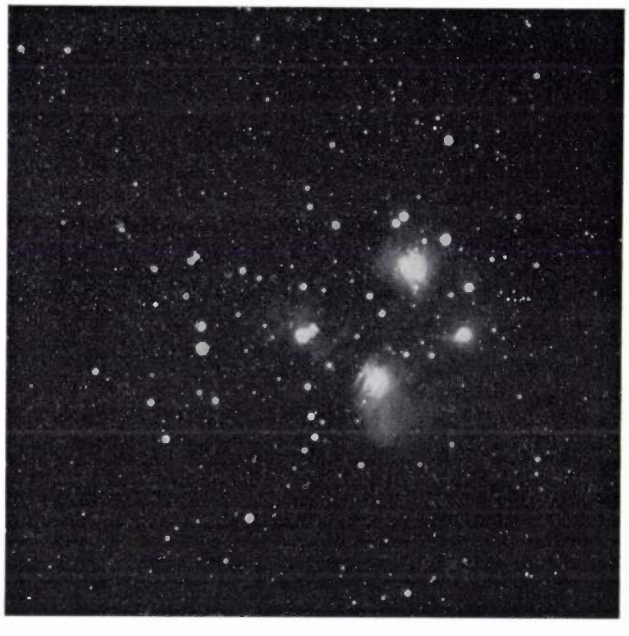


Fig. 2. An open cluster of stars, the Pleiades. This cluster will be observed primarily to define the main sequence in the ultraviolet, and further to study the reflection nebulae of various stars. These nebulae are clouds of dust and gas surrounding a star, which can be seen because they reflect the light from the star.

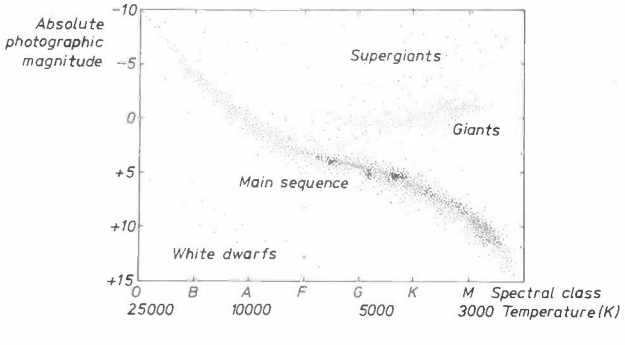


Fig. 1. A Hertzsprung-Russell diagram. The vertical scale gives a measure of the logarithm of the absolute luminosity $L(\lambda)$, here the absolute visual magnitude M_{pg}, and the horizontal scale gives a measure of log T_e, here both the spectral class and the temperature T_e itself. In a diagram of this type nearly all the stars lie in a narrow strip, forming the 'main sequence'. Other stellar sequences include the giants, white dwarfs, etc. To each of these sequences a stage of evolution in the life of the stars can be assigned. The number of stars in each sequence gives an indication of the time spent in that stage. It is evident that the stars spend most of the time in the stage represented in the main sequence.

[2] A general description of the ANS project is given in: W. Bloemendal and C. Kramer, Philips tech. Rev. 33, 117, 1973 (No. 5). The initials ANS form an acronym for *Astronomische Nederlandse Satelliet*.

pendicular to the galactic plane, shows that these objects lie in this band. Presumably these stars belong to the oldest in our own galaxy (known as population II stars) and are in the last stage of their evolution. They are stars of roughly one solar mass. Their content of elements heavier than hydrogen and helium — often incorrectly called 'metals' by astronomers — is much lower than that of the Sun. The metal content is another factor, not mentioned previously, that affects the location of a star in the Hertzsprung-Russell diagram. The He content is also of significance here. In the far-evolved population II stars the hydrogen has been largely converted into helium by nuclear processes. Through the combined effect of a low metal content and a high helium content a star of about one solar mass is hotter at the end of its life than the Sun (which has a temperature of about 5700 K) and is somewhat brighter.

By studying these hot subdwarfs in the ultraviolet we hope to be able to classify them better in terms of temperature and luminosity, and by comparing the results with theoretical calculations we may perhaps achieve a better understanding of the later stages of evolution.

A serious difficulty encountered in studying hot stars in the ultraviolet is that the space between the stars contains very small particles of dust that scatter and absorb the light from a star. This has two effects: it attenuates the starlight, and it alters the wavelength distribution of the intensities because the interstellar dust scatters ultraviolet more strongly than visible radiation. The interstellar red shift can be corrected by using suitably selected linear combinations of the measured UV intensities. The attenuation, however, is an essential limitation, and it prevents us from studying the intrinsically brighter hot stars to such great distances as we should wish.

Fig. 3. A cluster of young hot stars, contained in an emission nebula. Observations of this cluster can teach us more about star formation and circumstellar matter.

On the other hand one of our objectives is to study this interstellar dust itself, for example in a region where stars may be in the process of formation, as in *fig. 3*.

Finally, we have also invited other interested astronomers to submit proposals for small observation programmes. Some twenty interesting proposals have already been received and accepted.

The UV instrument: requirements and design

In the design of an astronomical instrument to be carried by a man-made satellite many requirements have to be satisfied. A distinction may be made between primary and derived requirements. The latter category includes many that only became evident during the design phase, or even later, in the production or trial phases. They arise because of the solutions found for the primary requirements.

The primary requirements may be summarized in two groups. One includes the requirements set by the measurements to be performed, and the other includes those that arise from the nature of the satellite and from the fact that it will be in space during the measurements.

In our case the astronomical requirements are as follows:

— The instrument must be able to perform measurements in five broad bands in the ultraviolet, at 1550 Å, 1800 Å, 2200 Å, 2500 Å and 3295 Å.

— The spectrometer must have an entrance aperture corresponding to a field of view of 2.5×2.5 minutes of arc.

— The sensitivity must be such that measurements can be made on stars with a brightness down to the 10th magnitude.

— The dynamic range must be sufficient to permit the analysis of brighter stars as well, so that a comparison can be made with measurements performed using less sensitive instruments (experiments from the Earth or previous satellite experiments).

— The instrument must have room for a star tracker sensitive enough to permit attitude control to within an accuracy of approximately 1 minute of arc.

The designer's assignment was to meet these requirements without coming into conflict with the requirements of the second category:
— the weight is subject to an upper limit,
— the volume is limited,
— the instrument must be strong enough mechanically to survive the launch,
— the equipment must be capable of withstanding the expected thermal conditions.

To help meet the weight requirement, almost the entire frame of the instrument is made of a magnesium alloy (relative density 1.8). The weight cannot be reduced too far, because the instrument must have sufficient mechanical strength to be unaffected by the predicted vibration levels. At low frequencies these can reach about 14g. Furthermore, since the instrument is fairly long, and is attached at only one point to the satellite structure, a fairly high rigidity is required to prevent vibrational amplification in the instrument. This could have a particularly serious effect on the photomultiplier tubes and on the star-tracker tube.

Once the satellite is in orbit, the instrument will be in an environment of fluctuating temperature. There will also be temperature differences of between 10 and 20 degrees between the entrance aperture, which is directed at the cold outer space, and the interior parts. Temperature differences of this order had to be taken into account in the design to guarantee the stability of the optical system and also to ensure that the electronics and the moving parts would function properly.

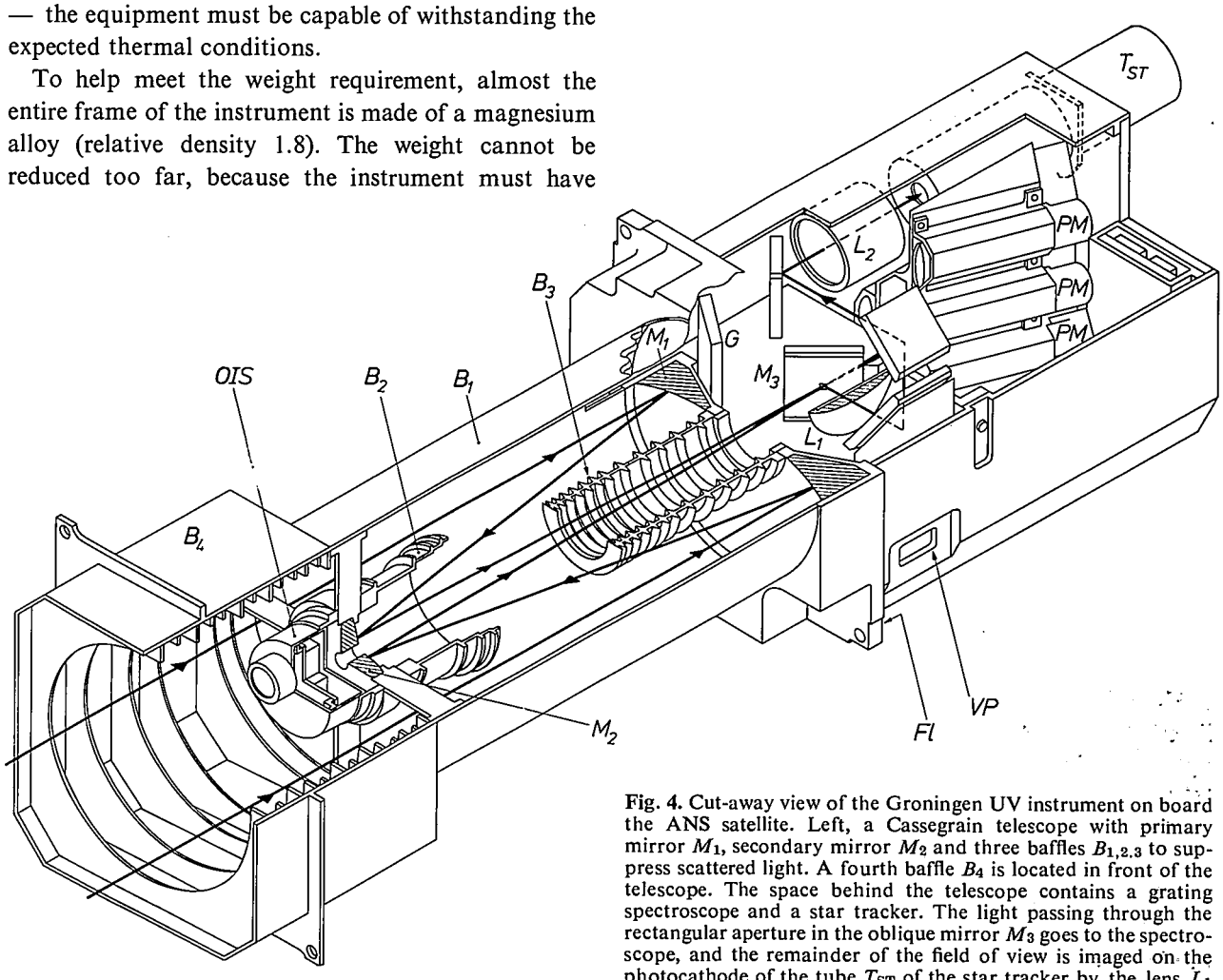


Fig. 4. Cut-away view of the Groningen UV instrument on board the ANS satellite. Left, a Cassegrain telescope with primary mirror M_1, secondary mirror M_2 and three baffles $B_{1,2,3}$ to suppress scattered light. A fourth baffle B_4 is located in front of the telescope. The space behind the telescope contains a grating spectroscope and a star tracker. The light passing through the rectangular aperture in the oblique mirror M_3 goes to the spectroscope, and the remainder of the field of view is imaged on the photocathode of the tube T_{ST} of the star tracker by the lens L_1 via a number of plane mirrors and the lens system L_2 (see the diagram of the path of the rays). G grating and PM three of the five photomultiplier tubes of the spectroscope. Apart from B_4, which is attached to the satellite frame separately, the telescope, spectroscope and star tracker form a single rigid assembly, fixed to the frame by the flange Fl. OIS overillumination sensor. VP window of the fine solar sensor, which is part of the attitude-control system [3].

Arrangement and special features of the instrument

The general arrangement of the instrument is illustrated in the cut-away view shown in *fig. 4*. The input of the system is a reflecting telescope (mirrors M_1 and M_2). This is needed to collect a sufficiently large flux and an image of the starry sky with a reasonably good optical resolution.

[3] The attitude-control system of the satellite is described in: P. van Otterloo, Philips tech. Rev. 33, 162, 1973 (No. 6). A more detailed description of the sensors will be given by W. J. Christis, P. van Dijk and A. J. Smets in Philips tech. Rev. 34, 1974 (No. 8).

The light from the object star passes through a square aperture in the mirror M_3 to the spectrometer. The remainder of the field of view is imaged by the lens L_1 on to the star tracker, which, together with the attitude-control system of the satellite [3], keeps the telescope pointed at the object. The spectrometer and the star tracker are both located in the right-hand part of the housing.

A grating G resolves the light entering the spectrometer into its spectral components, which are then separately measured by five detectors. The figure shows three of the five photomultiplier tubes PM, which serve as detectors.

The separation by wavelength can of course also be done with filters, but our system offers three advantages. Firstly, the five bands are measured simultaneously, secondly, the transmission is high (30 to 40%) and thirdly, the transmission characteristic in the individual bands is virtually rectangular (*fig. 5*); this kind of characteristic can only be obtained with filters at the expense of a considerable transmission loss.

The five detectors are completely independent: each detector has its own power supply, amplifier, discriminator and counting register. There is consequently virtually no risk of crosstalk or interference between the individual channels.

The discriminators and counting registers are not located in the instrument itself but elsewhere in the satellite, in a separate unit that forms the electronic link between the instrument and the onboard computer [4] and telemetry. The control of the instrument is also effected through this unit.

The number of pulses recorded in the registers is directly proportional to the received flux, after subtracting the background, of course. The background consists of thermal noise from the photodetectors and of high-energy particles from the Van Allen belts.

At high flux values the proportionality no longer holds. When the contribution from the Van Allen belts is small, the dynamic range is about 10^5.

The photometric stability, i.e. the constancy of the sensitivity, is at an optimum because the photodetectors operate in the pulse-counting mode, i.e. they count the separate photoelectrons. This differs from the d.c. mode, which is more dependent on the stability of the high-voltage supply and of the amplifiers. Nevertheless the sensitivity is certainly a function of temperature, and also to some extent a function of time. To measure these variations in sensitivity a Čerenkov source is incorporated in the instrument as a calibration source. Its operation is based on the effect in which particles travelling through a medium at a speed greater than the velocity of light in that medium ($v > c/n$) produce light (the Čerenkov effect). In the usual construction, which is used here, a substance that emits beta particles (e.g. ^{90}Sr) is applied to a quartz or BaF_2 substrate. The Čerenkov source is mounted in a shutter that has three positions: 'open', 'closed' and 'calibration'. In the calibration position the source is moved to the place where the starlight enters the spectrometer in the 'open' position. In this way it is possible to check the sensitivity of the spectrometer in all channels while in orbit by a command from Earth or from the onboard computer.

This method of calibration does not of course take into account the effect of the telescope on the sensitivity of the instrument (the 'telescope degradation'). For a true calibration we are therefore dependent on well known and frequently visible stars.

As fig. 4 shows, in addition to the telescope and the spectrometer section, the instrument includes a light baffle B_4 in front of the telescope to keep out scattered light.

The telescope and the spectrometer housing are mounted on a large rectangular flange *Fl*, which attaches the whole assembly to the satellite frame. For

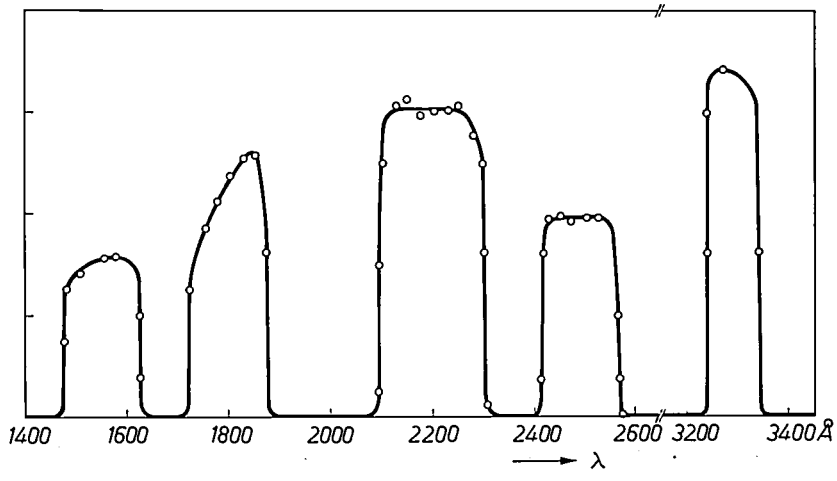


Fig. 5. Transmission characteristic of the UV spectrometer. The response is plotted (in arbitrary units) as a function of the wavelength λ.

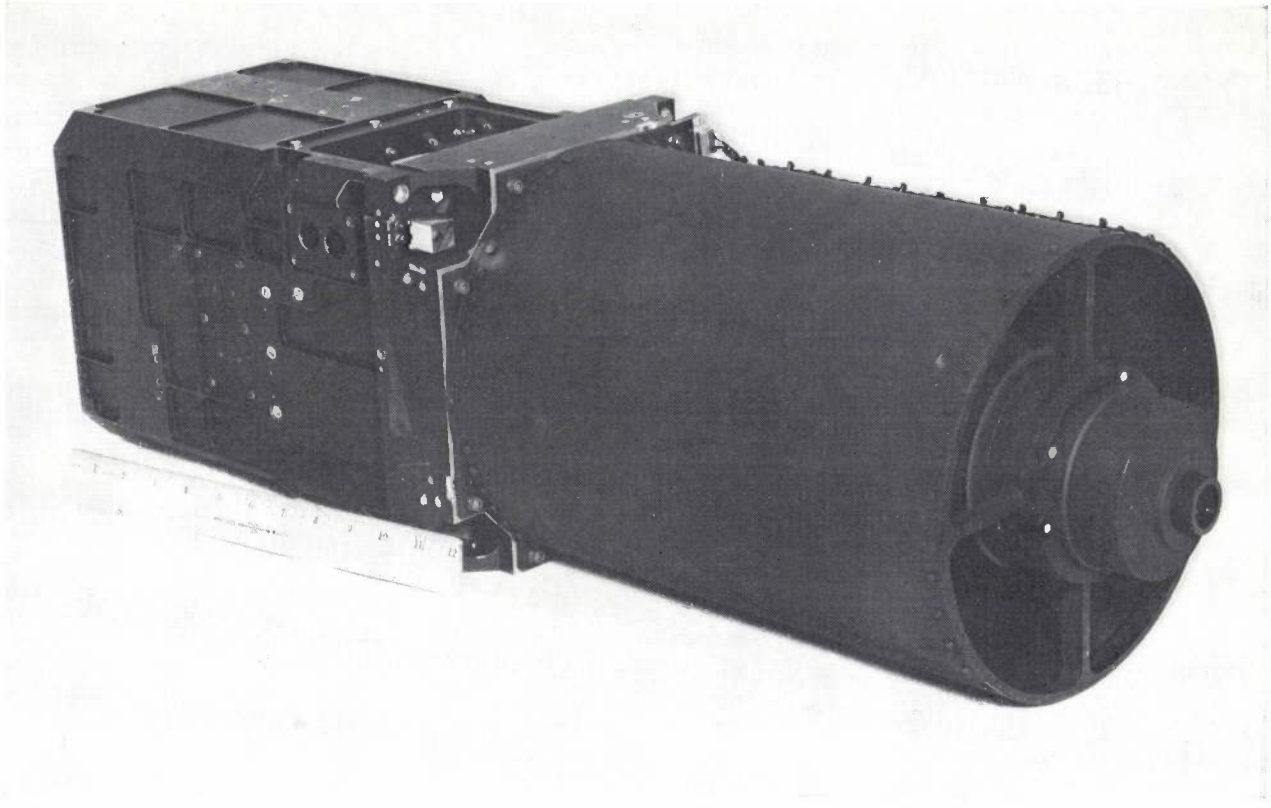


Fig. 6. The Groningen UV instrument. *Right*, the telescope; *left*, the unit containing the spectroscope and the star tracker. The two openings at the top of the side wall are windows for the fine solar sensor (fig. 4, *VP*). The prisms on their right are used for alignment. (The ruler is marked in *inches*.)

mechanical and thermal reasons the light baffle B_4 is fixed to the satellite separately.

The primary mirror M_1 is mounted in a flange that keeps it rigidly connected to the satellite. The secondary mirror M_2 is fixed to the telescope tube by means of a rigid four-armed spider. The telescope tube is made of invar, which gives it the required optical stability in the operational temperature range.

Located on the outside of the secondary mirror is the over-illumination sensor *OIS*, a sensor with a wide field of view (semi-aperture angle 35°). When excessively bright objects, such as the Sun or the illuminated Earth, threaten to appear in the field of view of the telescope (semi-aperture angle 0.75°) this detector delivers a signal that closes a shutter in the light path to the star tracker and the spectrometer.

The instrument itself and a small part of its interior are shown in *figs. 6* and *7*.

Optical system of the star tracker

We have just seen that the telescope serves at the same time as a part of the spectroscope and as a part of the star tracker. As will be explained below, the parameters of the telescope have mainly been chosen to give the size of field and image quality required for the star tracker. The sensitivity of the star tracker is determined by the light-collecting surface of the telescope, the transmission of the optical system and the sensitivity of the photocathode. Our instrument is designed in such a way that the star tracker can observe stars down to the 8th magnitude. The field of view is determined by the requirement that there should be *two* guide stars for the determination of a celestial object. With the size of field we have chosen, 1.5×1.5 degrees, the chance that in most areas of the sky the field of view will include two stars that are brighter than the 8th magnitude is greater than 70%.

The correct imaging of such a large field on the photocathode requires a fairly complex optical system. This is why we have adopted a Cassegrain telescope of the Ritchey-Chrétien type, which has aspherical surfaces whose focused image is free from coma and spherical aberration.

The focused image of the telescope (diameter 40 mm) is projected as a reduced image on the photocathode of the star tracker, which has a diameter of only 15 mm.

[4] A description of the onboard computer is given in: G. J. A. Arink, Philips tech. Rev. **34**, 1, 1974 (No. 1).

The necessary reduction (0.275) could only be obtained in the available space in the spectrometer housing by employing a complicated imaging-lens system. The object distance required to produce such a large reduction was much larger than the housing of the instrument. This made it necessary to 'fold' the light path four times by means of four plane mirrors. We could of course have made the focus image of the telescope smaller, but then the smaller f/D value required would have increased the difficulty with scattered light, shortly to be discussed.

The image quality finally obtained and the optical parameters are presented in *Tables II* and *III*.

cathodes of the multiplier tubes are not equally sensitive at all points. If the field lenses were omitted and the tubes located directly behind the slits, the result would be an impermissible variation in the response during the movement of the object star in the entrance aperture. Since the angle of incidence of the light on the grating changes slightly when such a shift takes place, it does give rise to a spectral shift of about 20 Å, but this is only 20% of the bandwidth at the most.

The efficiency of the grating is of course greatest at the blaze wavelength, where it is 42%; in the other bands it is less, but nowhere is it less than 30%. The lens material and the material of the photocathode

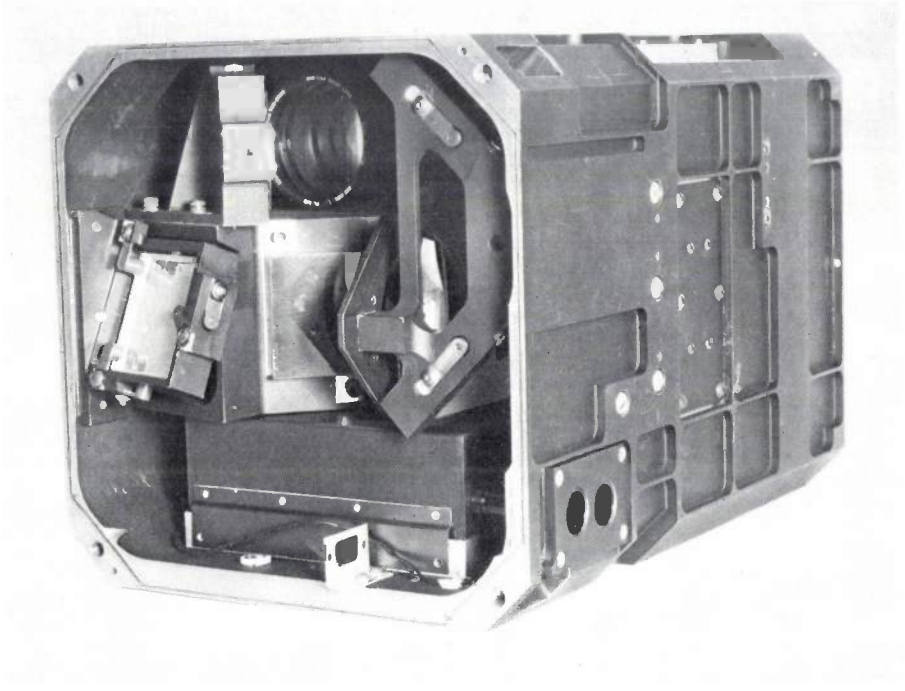


Fig. 7. As fig. 6, but now without the telescope and in approximately the same position as in fig. 4. Note the oblique mirror M_3 (fig. 4) with the lens L_1 on its right; the grating is on the left and the lens L_2 above it.

The optical system of the spectrometer

The spectrometer is of the Wadsworth type, as illustrated in *fig. 8*. The entrance aperture S_0 is situated at the focal plane of a concave spherical mirror M_4, which projects the incoming light as a parallel beam on to a similarly concave spherical refraction grating G. The spectral components emerging from the grating are thus focused again at a curved surface. In our case there are five slits S in this surface, and their position and width determine the passbands of the spectrometer. The transmitted light reaches the photomultipliers through small field lenses, which produce the image of the telescope aperture on the photocathode.

These field lenses are necessary because the photo-

were chosen so as to avoid detection of second-order light, in particular of the Lyman α line of hydrogen (1216 Å). The intensity of this light is considerable in the upper layers of the Earth's atmosphere. Most of the light leaving the grating is of the zeroth order, and this light is suppressed by a baffle constructed for the purpose. Measurements of the optical crosstalk between the different bands show that it is less than 10^{-3}.

Suppression of scattered light

In designing the instrument we paid careful attention to the requirements for the suppression of scattered light. Since the satellite will always be approximately

Table II. Principal data for the telescope

Type	Ritchey-Chrétien
Effective focal length	1500 mm
Effective aperture	*f*/6.6
Field of view	90×90 min. of arc
Axial resolving power	10 sec. of arc max.
Resolving power off-axis	see Table III
Transmission	65% (1400 Å < λ < 3300 Å) 70% (4000 Å < λ < 6000 Å)
Primary mirror (M_1)	225 mm
Focal length	500 mm
Secondary mirror (M_2)	92 mm
Distance M_1-M_2	350 mm
Distance from M_1 to image	100 mm
Light-collecting surface	266 cm²

Table III. Principal characteristics of the star-tracker optical system (the lenses L_1 and L_2 and the folding mirrors in fig. 4).

Effective focal length	80.26 mm
Magnification	0.274
Effective focal length of telescope and star tracker	411.5 mm
Effective aperture of star tracker	*f*/1.8
Size of image	10.8×10.8 mm (90×90 min. of arc)
Resolving power (90% energy threshold)	35 sec. of arc (diameter 0.07 mm)
Wavelength range	4000-6000 Å
Transmission, including telescope and folding mirrors	50%
Distortion	less than 1%

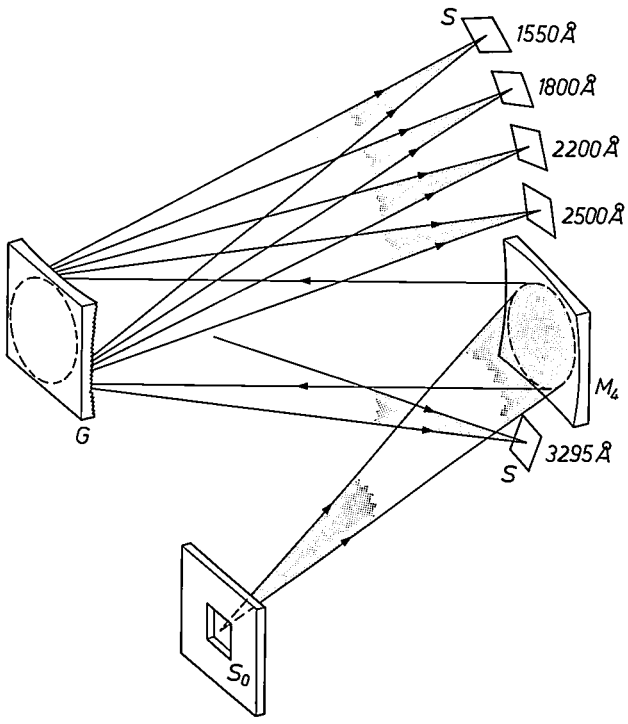


Fig. 8. Principle of the UV spectrometer (Wadsworth type). The spectral focusing is shown schematically as if the incident light contained only five discrete wavelengths. S_0 entrance aperture. M_4 collimator mirror. G grating. The exit slits of each of the five detectors are shown on the right, with an indication of the central wavelength of the transmitted band (see fig. 4).

above the terminator, it will constantly be illuminated by both the Sun and the sunlit Earth. Because of the changing attitude of the viewing direction in relation to the Earth the incident quantity of light will therefore not be constant. Furthermore the albedo of the Earth shows considerable local variations; above the polar caps it is about 0.75, but the average value is about 0.35.

The most unfavourable angle with the telescope axis at which the earthlight can enter is 36°. If the angle of incidence is less than this, the overillumination sensor will intervene and close the shutters.

The baffle system reduces the undesired light by such a large factor — greater than 6×10^7 — that even in the worst situation the permissible background for the star tracker is not exceeded. This is defined as the intensity per 45×45 seconds of arc of a star of the 10th magnitude, which requires a reduction factor of 2×10^7.

Some theoretical calculations were done on the design of the baffle system, but it proved impossible to make exact 'front-to-end' calculations for the whole system. The design is thus a 'best effort', based on requirements derived from partial calculations and relating to optimum geometry, sharpness of edges and surface coatings.

The rings in the cone are ground razor-sharp and their relative position is such that obliquely incident light can never reach the central aperture of the primary mirror. The front baffle is undoubtedly the most important component. The great lateral depth made it possible to limit the number of rings. The upper face of this baffle is chamfered to prevent sunlight entering at 90° from striking the inside wall. The inside surface of the telescope tube is covered with matt black paint ('Cat-O-lac').

The first design of the instrument was followed up with what was called an 'optical-calibration model', which fully represented the optical aspects. It was used to test the different baffle designs. A simulation arrangement was built in which the sunlit Earth was represented by a 90° sector of a circle, diffusely illuminated from behind. The instrument was set up opposite to it at the worst angle likely to be encountered in operation. The whole arrangement was surrounded by a black velvet 'tent' to simulate the dark sky. The tent was of course also lit by the artificial Earth, and consequently the luminous flux in the field of view of the telescope was not infinitely small. The result was an upper limit of about 6×10^7 to the measurable reduction factor.

Calibration of the spectrometer

In general the response of a spectrometer is given by

$$R = S(\lambda)F(\lambda),$$

where $S(\lambda)$ is the sensitivity in, for example, counts photons^{-1} cm^2 and $F(\lambda)$ is the incident flux in photons cm^{-2} s^{-1}.

To study the spectra of stars it is very important to know the variation of $S(\lambda)$. A second aim is to relate $S(\lambda)$ to a standard, to obtain an absolute measure of the observed fluxes. For this purpose a calibration system has been built at Groningen, which can direct a parallel beam on to the telescope by means of a 30-cm collimator. The collimator and the instrument are both contained in a vacuum chamber to eliminate absorption losses in air, which can be considerable below 2200 Å. The light in this beam comes from a monochromator, which is continuously adjustable over the entire range of wavelengths.

The calibration amounts to determining $F(\lambda)$ as accurately as possible. Unfortunately the beam intensity is not constant over the cross-section, which means that the intensity distribution has to be measured. Instead of $F(\lambda)$ we therefore take $\int F(\lambda)\mathrm{d}O$ divided by the effective area of the primary mirror.

The detector used for this has a window of terphenyl, which gives it an approximately constant sensitivity in this range of wavelengths. The sensitivity was subsequently compared with that of two detectors from the University of Wisconsin, which had been calibrated against synchrotron radiation. The intensity of synchrotron radiation is theoretically well known and is therefore used as a standard UV source.

The same calibration system was used for determining such characteristics as optical crosstalk, the exact location of the bands, the spectral resolving power, the sensitivity to white light and the nonlinearity at the limit of the dynamic range.

Summary. The Groningen astronomical instrument on board the Netherlands astronomical satellite (ANS) is a grating spectrometer capable of simultaneous measurements in the ultraviolet in five wavelength bands at 1550 Å, 1800 Å, 2200 Å, 2500 Å and 3295 Å. Since measurements can also be made on faint stars (down to the 10th magnitude) the instrument is very suitable for an improvement of the criteria used in the classification of stars. The field of view is kept to only 2.5×2.5 minutes of arc to reduce the contribution from the background and to allow measurements on closely adjacent stars. The light is gathered with a Cassegrain telescope fitted with four baffles to avoid scattered light. Part of the light-passes through an aperture in one of the mirrors to the star tracker that forms part of the attitude-control system.

Observation of cosmic X-ray sources with the Netherlands astronomical satellite (ANS)

A. C. Brinkman, J. Heise and C. de Jager

Introduction

The history of X-ray astronomy is short. It began on 12th June 1962 when a rocket was launched from the U.S.A. with the object of determining whether the Moon, which at that time was in the southern sky, emitted X-radiation. It was indeed observed that, whenever the detector in the axially rotating rocket was pointed towards the south, the X-ray flux showed an increase. Although the resolving power of the detector was poor, it could nevertheless be seen that the peak of the radiation did *not* coincide with the direction of the Moon, but to a direction about twenty degrees away from it (*fig. 1*). The source of this radiation appeared to be in the constellation of Scorpius; it was later called Sco X-1, the first X-ray source in the Scorpius constellation.

On the same occasion it was found that a diffuse background of X-radiation was present that could not have been caused by extraneous particles. Thus, two fundamental discoveries were made at the same time: the background radiation and an X-ray source.

In subsequent experiments carried out during the sixties with rockets and also with balloons, the number of known sources was steadily increased, and more exact information was also obtained about the diffuse background. Since the time in which observations can be made with a rocket is short, owing to the short flight duration, the threshold sensitivity was not particularly good in spite of the use of very large detectors, and varied between 1 and 0.1 photon/cm^{2}s.

The launching off the coast of Kenya of the satellite UHURU (Swahili for 'freedom') on 12th December 1970 was an immense step forward. This satellite, which had a long and useful life, had detectors in its equatorial plane which repeatedly scanned the same strip of the sky.

These facilities made it possible to reduce the weakest observable photon flux to about 10^{-2} photon/cm^{2}s, and also — and this proved to be very important — to observe a number of sources for periods ranging from days to weeks. This revealed the presence of sources that show very considerable periodic variations in strength. Some of these sources were subsequently identified as X-ray-optical binary stars, i.e. two stars that rotate around each other in Keplerian orbits, so that part of the X-radiation is periodically eclipsed as seen from the Earth. It looks as if the observation of one of these objects provided the first clear indications of the existence of a 'black hole' in the galactic system. The variation in intensity was also investigated in detail and resulted in speculations about the evolution of neutron stars. This is especially interesting when the neutron star forms part of a binary system. Furthermore, in some X-ray sources an irregular and fairly slow variation was discovered.

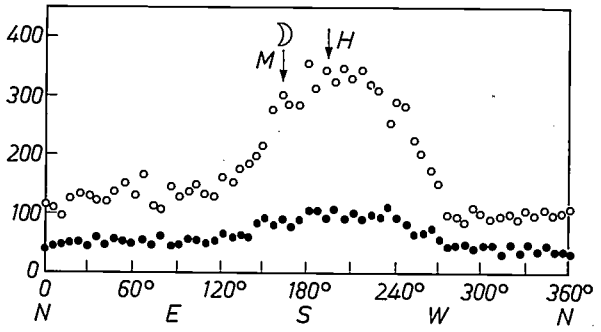


Fig. 1. 12th June 1962: discovery of the X-ray source Sco X-1 and of diffuse radiation [1] [2]. *M* position of the Moon. *H* position of the magnetic south pole.

It is scarcely surprising that the fascinating results achieved in this new area of research within such a short time gave rise to several plans to launch X-ray satellites. A survey of all the X-ray satellites that have meanwhile been launched, and those in preparation or under consideration, will be found in *Table I*.

The feasibility of carrying out observations on cosmic X-ray sources is limited by two effects. The first is the absorption of X-rays by the Earth's atmosphere (see *fig. 2*), which is why these observations have to be made with rockets or satellites. It is true that the hard X-radiation from outer space can also be measured

Dr Ir A. C. Brinkman is a senior scientific officer of the Space Research Laboratory of the Institute of Astronomy at Utrecht University; he is the Local Project Manager for the ANS project. Drs. J. Heise is a scientific officer of the Space Research Laboratory; Prof. Dr C. de Jager is a Professor of Space Research at Utrecht University.

[1] R. Giacconi, H. Gursky, F. R. Paolini and B. B. Rossi, Phys. Rev. Letters 9, 439, 1962.
[2] R. Giacconi and H. Gursky, Space Sci. Rev. 4, 151, 1965.

Table I. X-ray satellites launched, in preparation or under consideration

Satellite	Instrument	Operating mode	Energy range	Launching	Status	Country (Organization)
SAS-A (UHURU)	propl counters	scanning	2- 10 keV	Dec. 1970	in operation	U.S.A.
OSO VII	propl counters scintillator	rotating rotating	1- 60 keV 7- 300 keV	Nov. 1971	in operation	U.S.A.
OAO-C (Copernicus)	parab. ref. propl counters	pointing	0.2- 10 keV	Aug. 1971	in operation	U.S.A.
UK-5	propl counters Bragg crystal spectrometer	pointing scanning	0.3-2000 keV	1974	in prepn	U.K.
ANS	propl counters parab. ref. Bragg crystal spectrometer	pointing scanning	0.2- 40 keV	Aug. 1974	in prepn	NL
SAS-C	propl counters modulation coll. parab. ref.	pointing scanning	0.2- 70 keV	1975	in prepn	U.S.A.
OSO-I	propl counters scintillator	pointing scanning	0.2-1000 keV	1975	in prepn	U.S.A.
CORSA	propl counters scintillator	pointing scanning	0.2- 100 keV	1975	in prepn	Japan
UK-6	parab. ref.	pointing	0.2- 10 keV	1975	in prepn	U.K.
HEAO-A	propl counters modulation coll. crystal spectr.	scanning pointing	0.1- 150 keV	1978	in prepn	U.S.A.
EXO	imaging telescope	pointing	0.05- 40 keV	1979	under considn	U.K.,NL, W. Germany
EXOSAT	parab. telescope	pointing	0.1-20 keV	1979	in prepn	ESRO
HEAO-B	imaging telescope	pointing	0.1-1.5 keV	1979	in prepn	U.S.A.

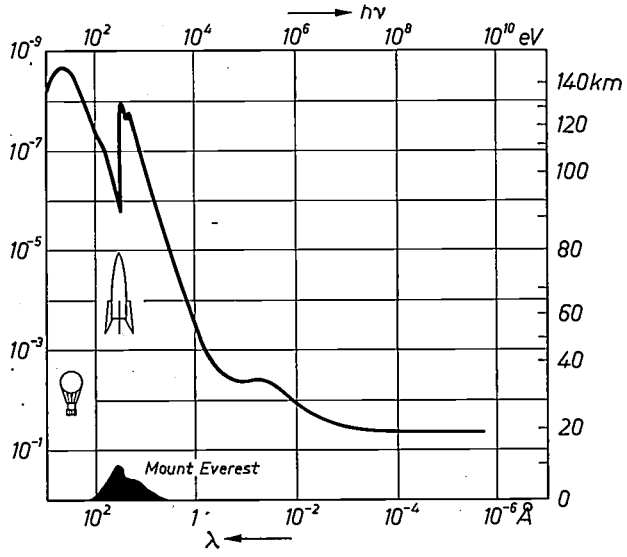


Fig. 2. Absorption spectrum of the Earth's atmosphere. The line gives the height at which the transmission of the corresponding wavelength is 0.5 [3].

with equipment carried by balloons — provided they reach a height of more than 40 km — but the measurements are seriously disturbed by the effects of cosmic-ray particles and their decay products in the Earth's atmosphere. If only for this reason, the observation instrument should preferably be sent to greater altitudes. This is even more important in measuring the softer X-rays, which are in a spectral region that has scarcely been investigated as yet and which we hope to learn more about with the ANS experiment. Satellite observations are also essential to permit a sufficient number of photons to be collected and to enable the variability of various X-ray sources to be studied.

The other limitation is due to interstellar gas. This consists mainly of hydrogen, but it also contains carbon, nitrogen, oxygen, neon and other gases. The absorption caused by the interstellar gas is wavelength-dependent (*fig. 3*). Although the density of the hydrogen particles in interstellar space is only of the order of 1 particle per cubic centimetre, and that of the other substances is smaller still, the strong continuous Lyman absorption of hydrogen is in itself sufficient to limit interstellar visibility at a wavelength of 300 Å to a few light-years. It is only at wavelengths of about 10 Å or shorter that we can penetrate to the galactic centre. This means that all we can see in the as yet virtually unexplored region of long X-ray wavelengths are the nearby X-ray sources. The more remote sources will be seen in a kind of haze, like a lantern in the mist.

In the next section we shall first go into some current

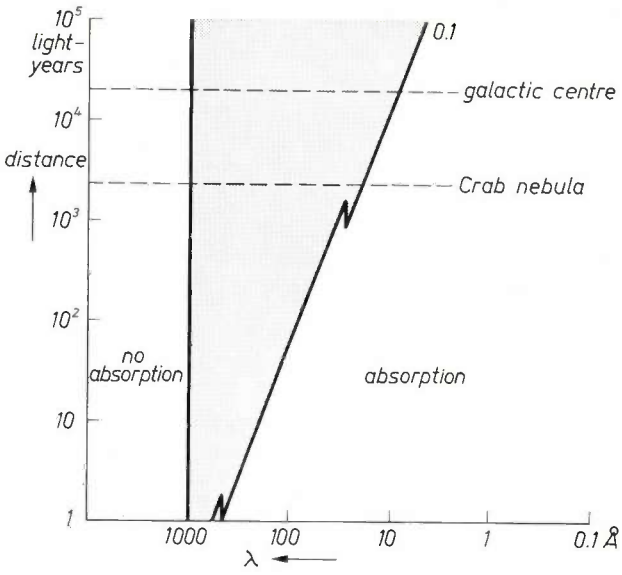


Fig. 3. Hypothetical absorption spectrum of interstellar gas assumed to consist of one hydrogen particle per cubic centimetre. The lines indicate the distance (in light-years) over which the absorption of the radiation is 0.1. Most of the radiation from the Crab nebula, which is about 2000 light-years away, will be transmitted through at about 14 Å; at 100 Å the nebula is invisible. Interstellar 'visibility' is apparently limited in the very soft X-ray range; at 100 Å not much farther than 100 light-years can be seen.

problems of X-ray astronomy a little more deeply and consider the contribution which the ANS experiments might make towards their solution [4]. We shall then discuss the construction and operation of our experimental onboard equipment for measuring soft X-radiation. In the final section we shall briefly describe the American equipment for measuring hard X-radiation [*].

X-ray astronomy

The enormous advance of X-ray astronomy in only ten years has demonstrated the capability of this branch of astronomy to provide important new information not only on the nature of various interesting astronomical objects but also on the structure and composition of interstellar matter, and perhaps even of intergalactic matter.

Quite a number of X-ray sources that have been discovered in the past ten years have totally unexpected characteristics. Some of these objects emit tens of thousands of times more energy in the X-ray region than in all the other parts of the electromagnetic spectrum taken together (optical, infrared, radio, etc.). The diffuse background of X-radiation extends over a very wide range of energies, from soft X-radiation with wavelengths of about 50 Å up to the gamma-ray range [5]. A better understanding of this diffuse X-radiation could make an important contribution towards the solution of problems concerned with the structure and evolution of the universe.

One of the first important contributions to the theory of stellar X-ray sources was the discovery that various sources are located in a binary system and are associated with very compact stars, in the first place with neutron stars but also perhaps with 'black holes'. As a result a number of fundamental quantities are now known, such as the masses and dimensions of the X-ray sources, and in some cases the distance of the system from the Earth. Neutron stars provide a unique field of research for high-energy physicists. The densities in neutron stars (10^{14}-10^{15} g/cm^3) are many orders of magnitude greater than could ever conceivably be reproduced under laboratory conditions. The gravitational fields are so strong ($g = 10^{13}$ cm/s^2 at the surface) that Newton's laws are no longer applicable, and consequently the relativistic theories of gravitation (including the general theory of relativity) can in principle be tested against the observations. This applies in particular to one of the most important theoretical predictions of relativistic theories of gravitation: the existence of the black holes we have just referred to. These are objects whose radius R is smaller than the gravitational radius (or Schwarzschild radius) R_g. For a non-rotating black hole $R_g = 2\,GM/c^2$, where M is the mass of the object, G is the constant of gravitation, and c is the velocity of light. If M is equal to 1 solar mass, then R_g is about 2 km.

A characteristic of black holes is that neither matter nor electromagnetic radiation can escape from the region enclosed within the radius of gravitation, hence the name. Such objects can therefore only be observed by virtue of their gravitational field.

The X-ray source Cyg X-1 could well have some connection with a black hole. The X-ray emission probably originates from the immediate vicinity of the black hole, where the potential energy of the matter falling into the hole is partly radiated at a temperature of 10^7-10^8 K. The difference between the potential energy thus released from a neutron star and that from a black hole is not great enough, however, for us to be able to distinguish between them with any certainty, but we shall return to this subject presently.

[3] R. Giacconi, H. Gursky and L. P. van Speybroeck, Ann. Rev. Astron. Astrophys. 6, 373, 1968.

[4] A general description of the ANS project has been given by W. Bloemendal and C. Kramer in Philips tech. Rev. 33, 117, 1973.

[5] Relatively hard radiation is usually characterized by the quantum energy E, relatively soft radiation by the wavelength λ. The Netherlands satellite operates in the transitional region ($E = 0.1$-10 keV), in which both measures are used somewhat indiscriminately. The relation between them is: $E(\text{keV}) \times \lambda(\text{Å}) = 12.39$.

[*] The final section is taken from a text kindly placed at our disposal by Dr. H. Gursky, Smithsonian Astrophysical Observatory, Cambridge, Mass., U.S.A.

First of all we shall give a brief outline of the present situation in X-ray astronomy, as mainly determined by measurements in the energy range above 1 keV (wavelengths shorter than about 10 Å). In the next section we shall then look at the significance of the soft X-radiation (wavelengths longer than about 20 Å), the wavelength region with which the Utrecht X-ray experiment will largely be concerned.

Nature, number and spatial distribution of the X-ray sources

Some 160 X-ray sources are now known. They vary in relative intensity (the flux on Earth) from 100 photons/cm²s in the wavelength region of 1-10 Å (the brightest source Scorpius X-1) to 0.01 photon/cm²s (the detection limit of the UHURU instruments). The positions of these sources are not yet very accurately established. The margin of uncertainty is about 1 square minute of arc for the strongest source to a few square degrees for some of the weaker sources. This is a serious handicap in X-ray astronomy, since identification with optical or radio objects or both is therefore difficult, which makes it difficult to obtain a better understanding of the type of object and of its physical nature.

Their distribution in the sky (see *fig. 4*) differs from that of the stars. In the stellar distribution there is a marked concentration of faint stars towards the Milky Way (hence the name), whereas the X-ray sources show a concentration of brighter objects towards the plane of the Milky Way, with a more or less isotropic distribution of faint sources beyond it. This suggests a subdivision into *galactic* and *extragalactic sources*, a subdivision which is fully supported by those objects that can be optically identified. In addition, the almost complete absence of a stronger background radiation in the

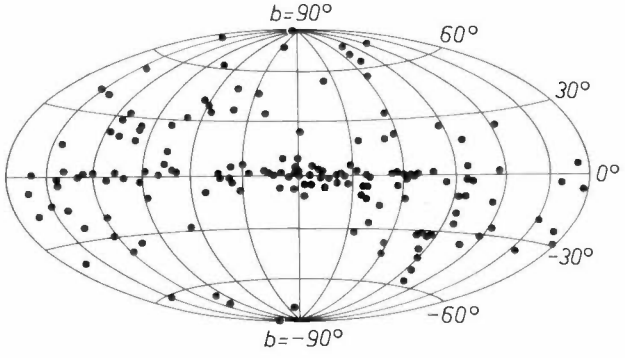


Fig. 4. Distribution over the sky of the 161 X-ray sources now known [6], in galactic coordinates. The galactic longitude is plotted horizontally, the latitude b vertically. Some sources are concentrated near the plane of the Milky Way ($b = 0$). These strong X-ray sources evidently belong to our own galaxy. There are also a number of X-ray sources, mostly weak, scattered more or less homogeneously over the sky, which have their origin in objects outside our galaxy (other galaxies, radio systems, quasars, etc.).

Fig. 5. The elliptic galactic system M87 is characterized by a stream of outflowing matter, which is the source of strong synchrotron radiation in the optical and radio bands. This source also emits X-radiation, but the resolution of instruments that have so far flown is too poor for determining the origin within this system.

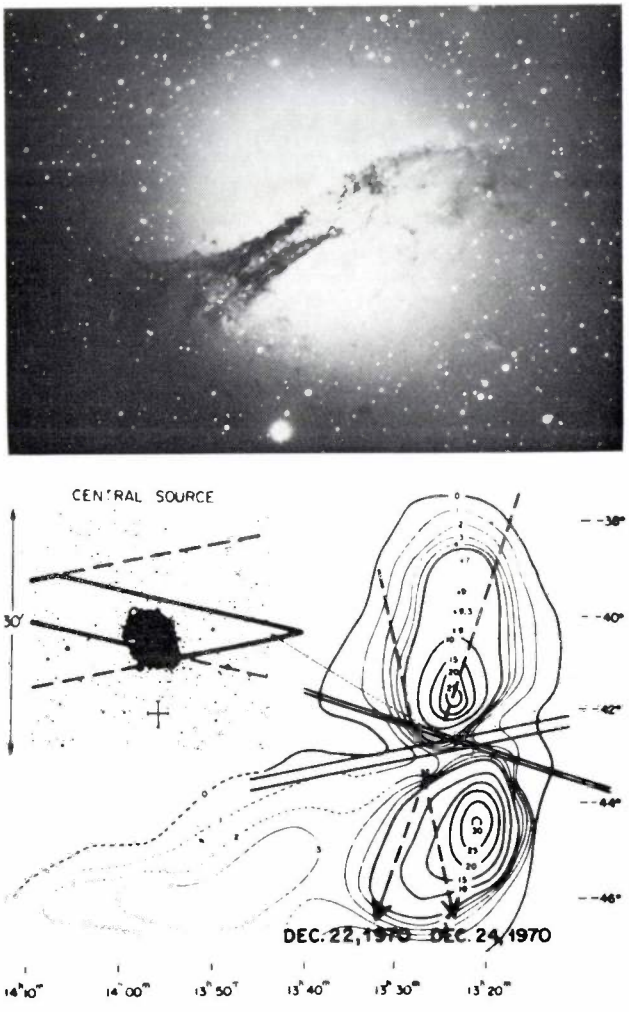


Fig. 6. The most powerful radio source in the Centaurus constellation, Cen A, the galactic system NGC 5128 has also been found to be a source of intense X-radiation. The figure shows lines of equal intensity in the radio band. The radio emission evidently originates from two regions that are very remote but are symmetrical with respect to the galaxy. The position of the X-ray source was determined by comparing observations obtained with a slit-shaped field of view. The X-ray source does not coincide with the radio source but with the galaxy itself. The inset shows a magnification of the central part of the figure, giving the bearings used; and a schematic diagram of the optical image of the system. Also visible in the photograph, even more magnified, is the optical image itself. The X-ray source may perhaps indicate the site of an enormous explosion, the radio source being the remnants.

galactic plane, which might be the result of the total radiation of a large number of sources that are spatially unresolved, indicates that we can in fact observe all the stronger X-ray sources in our own galaxy. The total number of strong X-ray sources in our galaxy is therefore surprisingly small, being no more than about 100.

Extragalactic X-ray sources

Various weak X-ray sources can be identified with extragalactic objects. One of the first extragalactic sources to be identified was the giant galactic system M87. This is the well known system (see *fig. 5*) from which a stream of matter escapes that forms the source of very strong synchrotron radiation in the optical and radio bands of the spectrum. The X-ray-emitting band extends in space over about 0.7° (this is a linear diameter of about 200 kpc [7]) and it is possible that the X-radiation is generated in a hot rarefied intergalactic gas between a number of members of the Virgo cluster of galaxies, one of which is the M87 system. (The whole Virgo cluster comprises some 1000 member and covers a region of about 10° diameter in the sky.) Other clusters of galaxies have also been identified as X-ray sources.

The existence of an intergalactic gas is tremendously important for our knowledge of extragalactic systems in general, and in particular for our knowledge of the structure of the universe. A rarefied gas between all galactic systems with an average density of approximately 10^{-5} atom/cm^3, for example, would be sufficient to bring the average density in the universe to about 10^{-29} g/cm^3, whereas observations on optically emitting matter (stars etc.) indicate an average density of only about 10^{-31} g/cm^3. The average density is an important parameter in models of the universe. Thus, if it were greater than 10^{-29} g/cm^3, it would mean that our universe was a 'closed' universe, in which the expansion now observed would gradually decrease.

Apart from clusters of galaxies, individual systems have also been identified as X-ray sources, such as the radio source Cen A (see *fig. 6*), the quasar 3C273 and various systems referred to as Seyfert galaxies, characterized by an intense activity in their nuclei. Our knowledge of the conditions prevailing in such galactic systems is still rather limited, however. X-ray astronomy can do much to add to that knowledge.

Finally, there is a fairly large class of extragalactic X-ray sources that have *not* been identified, even though their position is well enough known to make an optical identification possible. Plots of the numbers of X-ray sources as a function of their relative intensity show that their distribution is consistent with the assumption that the sources are homogeneously distributed in space, which means that they must be extragalactic. Further study of the characteristics of these intrinsic X-ray systems, about which we know very little as yet, is certainly needed.

Fig. 7. The Crab nebula, in the Taurus constellation, consists of the remnants of a star that exploded in the year 1054, a supernova. The nebula itself and the remnant of the exploded star, a neutron star, together constitute a source of intense X-radiation.

Fig. 8. The Veil nebulae in the Cygnus constellation. These are very old remnants of a supernova explosion that presumably took place 30 000 to 50 000 years ago. The temperatures in this nebula are of the order of a few million degrees Kelvin, and the nebula therefore emits only soft radiation in the X-ray band (wavelengths longer than about 10 Å) and is not detectable at shorter wavelengths. The UHURU satellite could not therefore observe this source, but it will be studied extensively with ANS.

Galactic X-ray sources

One of the first X-ray sources discovered was the Crab nebula, the still expanding remnants of the explosion of a supernova seen in the year 1054 (*fig. 7*). There are other, older supernova remnants that show X-ray emission, one of them being the gaseous nebula in the Cygnus constellation (Veil nebula, *fig. 8*). The sources extend in space from a few minutes of arc (Crab nebula) to a few degrees (Cygnus nebula).

Supernova remnants are among the best-explained X-ray sources, partly because of the wealth of informa-

[6] Taken from R. Giacconi, S. Murray, H. Gursky, E. Kellogg, E. Schreier and H. Tananbaum, The UHURU catalog of X-ray sources, Astrophys. J. **178**, 281-308, 1972.

[7] A pc (parsec) is 2×10^5 astronomical units (A.U.); 1 A.U. = 1.5×10^8 km.

tion obtained from the analysis of the emission in the optical and radio wavelength bands. The radiation from the Crab nebula is probably synchrotron radiation originating from the interaction between high-energy electrons with a magnetic field. Situated in the middle of the Crab nebula is a pulsar, a rapidly pulsating radio source (period 33 ms) whose optical and X-radiation also pulsates in intensity with the same period. It is thought that the pulsar is a rapidly rotating neutron star left over after the supernova explosion. It is the rotational energy of the pulsar that is finally emitted by the nebula in the form of continuous radiation. The period of the pulsar therefore gradually increases. The emission mechanism of the pulsar itself has not yet been explained in detail.

only be explained in terms of the rotation of *heavy* objects leads to the assumption, mentioned in the introduction, that X-ray sources of this kind must be neutron stars or perhaps black holes.

The development of the theory of X-ray sources received unexpected encouragement when it was found that some of them were members of binary star systems; the X-ray source is regularly eclipsed by an ordinary star. From the rotation period of such a system (about two days), from the duration of the eclipse (about half a day) and especially from the observed change in the pulsar period shown by the Doppler shift due to the orbital motion (see *fig. 11*), valuable information can be derived about quantities such as the mass of the X-ray source, the mass of the companion star, their

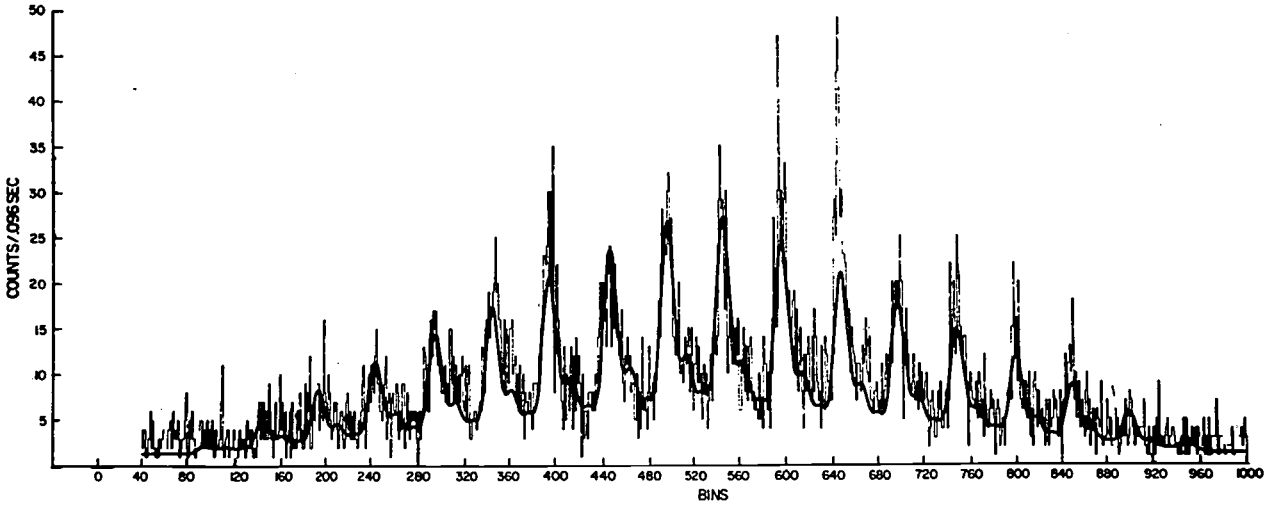


Fig. 9. The intensity of the emission from the X-ray sources Centaurus X-3 and Hercules X-1 fluctuates with periods of 4.8 s and 1.24 s respectively, hence the name X-ray pulsar. The pulsations are thought to be caused by matter falling inward on to the magnetic poles of a neutron star. (On neutron stars there is thus a kind of polar aurora, but of course very much more intense than those on Earth.) Because of the rotation of the neutron star these poles are obscured alternately as seen from Earth, which explains the pulsating nature of this X-ray source. The graph shows a measurement on Cen X-3, which extended over a time interval of about 100 seconds (1 scale division = 0.0965 s). [*]

X-ray sources in binary stars

Many X-ray sources exhibit fluctuations on every time scale on which they have so far been measured, from milliseconds to months. Two sources, Centaurus X-3 and Hercules X-1, are X-ray pulsars; they pulsate periodically with periods of 1.24 s and 4.8 s respectively (see *fig. 9*). There are a number of other sources that fluctuate irregularly; Cygnus X-1 and Circinus X-1 do so even within a few tens of milliseconds (see *fig. 10*). The inference from this is that an X-ray source of this type must have an extremely small diameter. The size of a region that flares up in say 10 ms must be smaller than the maximum distance that can be covered in that time by the information of this disturbance, i.e. smaller than 3000 km. The fact that on the other hand the very regular pulsations of the X-ray pulsars can dimensions and the distance between them. If the companion star has also been optically identified, something can usually be said as well about the distance to the system. Seven such binary-star X-ray source systems are now known.

The only system among these seven that has at the same time been optically identified, shows eclipses and is also an X-ray pulsar, is Hercules X-1. This is therefore the system on which we have the most information. In addition to the pulsar period of 1.24 s and the orbital period of 1.7 days, Her X-1 exhibits an unexplained on-off period of about 35 days (*fig. 12*); the source is seen to be alternately 'on' for 11 days and 'off' for 24 days.

[*] Diagram by courtesy of UHURU group, AS & E, Cambridge, Mass., U.S.A.

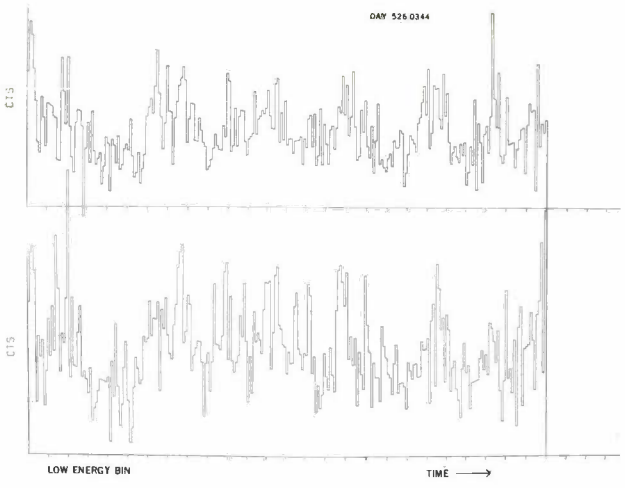


Fig. 10. The observed intensity of the emission from the X-ray sources Cygnus X-1 and Circinus X-1 shows irregular fluctuations on a characteristic time scale, which may be very short (milliseconds). The radiation source here could be the immediate vicinity of a 'black hole', an even more compact object than a neutron star. One scale division on the time axis is 0.0965 s. [*]

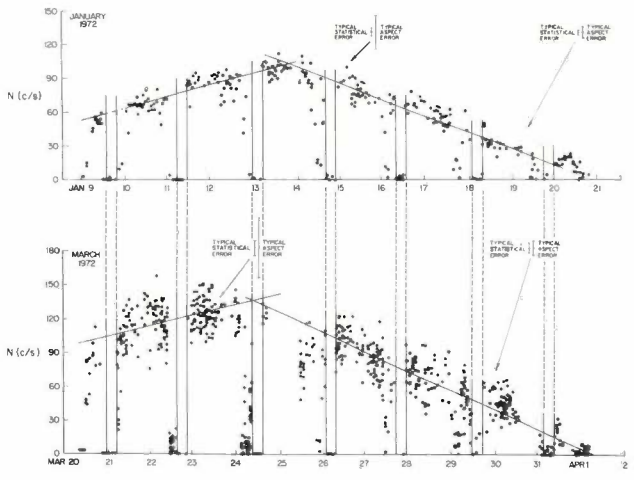


Fig. 12. The X-ray star Her X-1 shows an unexplained periodicity of about 35 days. The source is 'off' for about 24 days, and is then 'on' for the next 11 days, first flaring up very quickly and then slowly decaying. The figure shows measurements made during two such 'on' periods. During the 'on' period the intensity exhibits marked fluctuations, as observed from nearly all X-ray sources. [*]

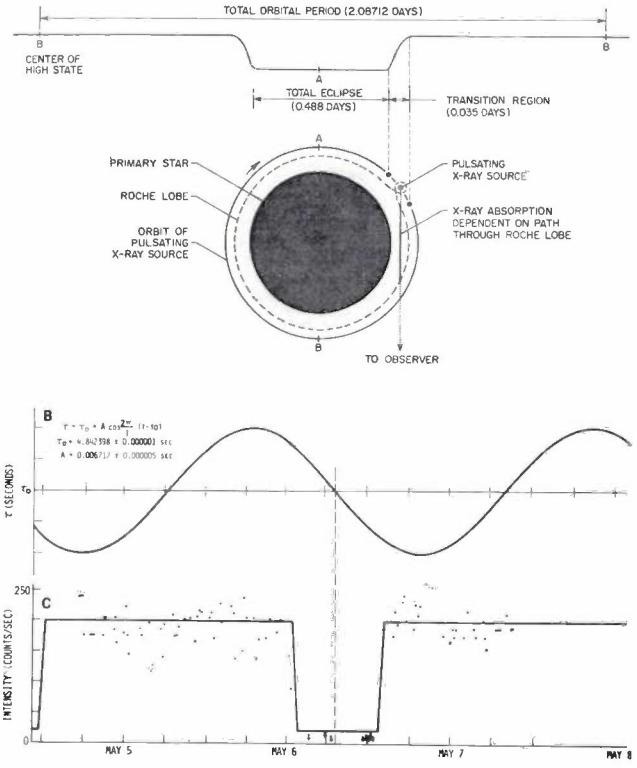


Fig. 11. *a*) In addition to the pulsar period some X-ray sources also show a periodic fluctuation due to the regular disappearance and reappearance of the X-ray source behind a large, heavy ordinary star. This indicates that such neutron stars are apparently a component of a binary stellar system. *b*) Measurements on an X-ray source of this type. The lower curve shows the variation in X-ray intensity due to the regular disappearance of the source behind the other star. Associated with the movement of the pulsar around the centre of gravity of the system there are also regular changes in the pulsar period as a result of the Doppler effect (upper curve). The Doppler shift makes it possible to measure radial velocities very accurately, and to determine together with the orbital period the mass of the star and of the neutron star. [*]

The total energy radiated by such an X-ray source lies in a fairly narrow band and amounts to 10^{36}-10^{38} erg/s, which is 10^3-10^5 times that from the Sun. This enormously high energy production by an object whose mass is of the order of one solar mass cannot be explained by nuclear fusion, as it can for ordinary stars. The energy production can however be understood in terms of the accretion of gas on compact objects, such as white dwarfs, neutron stars and black holes. In the case of a neutron star, for instance, with a mass equal to one solar mass and a radius of 10 km, some 10% of the rest-mass energy of the inflowing gas can be released in this way, i.e. 10^{20} erg/g. An accretion rate of 10^{16}-10^{18} g/s (which is about 10^{-10}-10^{-8} of a solar mass per year) would then be amply sufficient to explain the total energy radiation. This relatively small flux of matter may well originate from the companion star, for example on account of the gradual expansion of this star during its evolution, so that it transfers matter to the X-ray source (*fig. 13*). This hypothesis has been quantified by detailed accretion models.

The maximum radiation intensity of the X-ray source is that at which the radiation pressure of the source prevents further accretion. This maximum intensity, called the Eddington limit, is calculated on the basis of the equality between the acceleration of gravity and that of the radiation pressure. Let σ be the effective cross-section for Thomson scatter, and L the total intensity, then

$$\frac{\sigma L}{4\pi r^2} = \frac{GM}{r^2}; \qquad L_{\mathrm{Edd}} = 10^{38}(M/M_0)\ \mathrm{erg/s},$$

where M/M_0 is the mass of the X-ray source in units

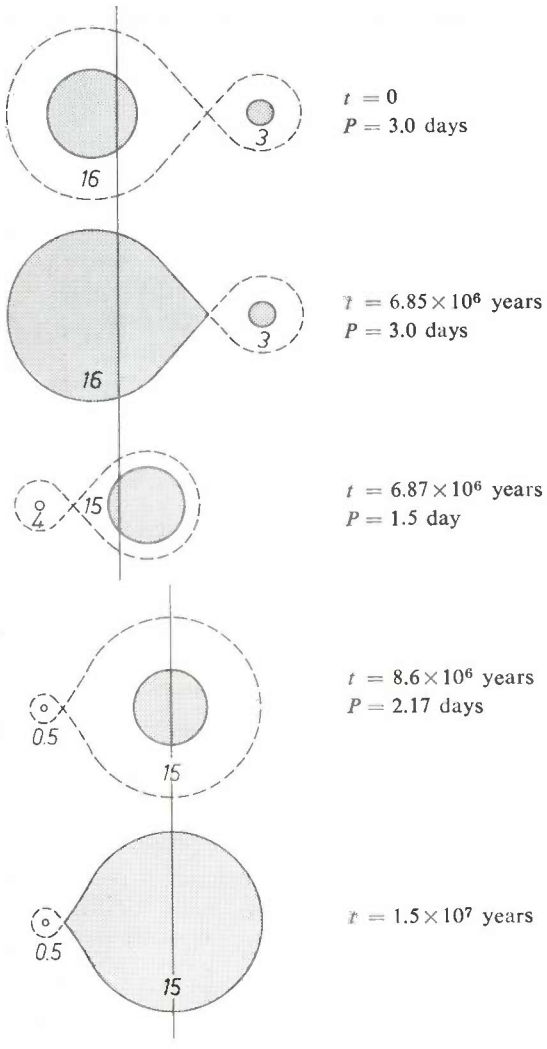


Fig. 13. Mass transfer in a binary star system may be an important factor in the evolution of the system. At the time $t = 0$ the binary consisted of two stars whose masses are equal to 16 and 3 solar masses. The rotation period P was 3.0 days. The dashed line is called the Lagrangian surface, being the equipotential surface in a revolving coordinate system that passes through the Lagrangian point. After 6.85 million years the more compact star, which evolves more rapidly, has become so large that it eventually fills a volume called the Roche lobe (the part within the Langrangian surface). From then on, mass transfer can take place from the heavier to the lighter star. As a result the rotation period of the pair changes. When the overflow has reached the equivalent of twelve solar masses, an almost pure helium star remains of four solar masses; the rotation period has then decreased to 1½ days. Theoretical investigations have indicated the probability that the helium star could then explode as a supernova, thus losing most of its mass. What remains is a neutron star of about 0.5 solar mass; the rotation period increases to more than 2 days. The originally lighter component which had meanwhile become the heavier of the two (15 solar masses) will thereupon evolve rapidly and after about 15 million years will fill its Roche volume, after which matter can again flow back to the old neutron star. The infall of matter on this star gives rise to the emission of X-rays, and the rapid axial rotation of the neutron star may even give rise to pulsar effects. The vertical lines indicate the position of the centre of gravity.

of one solar mass. This explains the observed maximum luminosity, which is of the order of 10^{38} erg/s.

The pulsating character of such an X-ray source is explained on the assumption that the neutron star has a strong magnetic field. The infalling matter forms two hot spots at the poles of this field, and as a result of the rotation of the star these hot spots cause the pulsed emission.

Black holes cannot possess a magnetic field, and therefore they will not be detectable as X-ray pulsars. In their case, irregularities in the accretion flux will be detectable as fluctuations whose time scale may be as short as that corresponding to the period of the most strongly bound stable orbit around a black hole. In a black hole of one solar mass this period is of the order of milliseconds. The main criterion for distinguishing between neutron stars and black holes, however, is the mass. The maximum possible mass of a neutron star is about 1.7 solar masses. Above that the pressure in the star can no longer compensate the force of gravity and the star, after a supernova instability, for example, will collapse and form a black hole. However, the calculated mass of the X-ray source Cyg X-1 which otherwise shows all the characteristics of a small compact object, is *five* solar masses. If this calculation is indeed correct, Cyg X-1 must therefore be a black hole.

Measurements with ANS in the soft X-ray region

The Utrecht X-ray experiment with the ANS satellite is primarily concentrated on soft X-radiation (wavelengths between 20 and 70 Å). No satellite experiments have been carried out previously in this energy range, and our knowledge is therefore based on a few sounding-rocket experiments, in each of which the observation time was only a few minutes. These observations showed that the sky observed in the soft X-ray band is quite different from the sky observed at wavelengths between 1 and 10 Å. Some sources, including old supernova remnants like the Veil nebulae in Cygnus (fig. 8), radiate intensely in the soft X-ray region, whereas they are barely observable if at all between 1 and 10 Å, the energy range of the UHURU instruments. Emission of this nature may be expected from cosmic plasmas with temperatures of a few million degrees Kelvin.

Apart from the emission of these new, typically 'soft' X-ray sources, that of the 'hard' X-ray sources in the soft band is also important. In other words, the extension of the spectrum up to 70 Å would provide valuable new information. This is because the effective cross-section for the photo-ionization of hydrogen, helium and other gases by X-radiation increases strongly with increasing wavelength (see fig. 3). Measurements of this absorption effect make it possible to determine the column density of these elements between observer and source. Particularly if the source is time-dependent, as in binary systems, such measurements yield unique

information on the distribution of gas around the system.

In the soft X-ray range a 'diffuse' background is found which is considerably more intense than would be expected from an extrapolation of the diffuse background between 1 and 10 Å. It is not yet clear whether this arises from a large number of spatially indistinguishable galactic soft X-ray sources, or whether the radiation is really diffuse and originates from interstellar matter. A satellite experiment in the soft X-ray band is necessary as a first step towards solving the above problems and towards explaining the phenomena observed. The longer observation time would permit greater sensitivity, while at the same time a better survey of the whole sky would be obtained.

Summarizing, we may say that the first aim of the observations is to measure spectra and intensity variations as a function of time for sources whose position in the celestial sphere is fairly accurately known. The second aim is to scan certain regions of the sky systematically, especially in the long-wave band, a spectral region that has not been explored in this way before. For this purpose the Netherlands astronomical satellite will use the 'slow-scan mode' in which the scanning speed is 0.4°/min [8].

The Utrecht instrument for measuring soft X-rays; design and characteristics

The requirements to be met by the observations together with the fact that the instruments are carried by a satellite which can be accurately pointed but has a limited memory capacity, a special orbit, a limited permissible weight and a limited power supply [4] [8] [9], had a number of consequences for the design of the experimental system.

It is not possible to design a detector that is sensitive over the whole range from 6 keV to 150 eV. The Utrecht experimental package therefore consists of two detection units together with an electronics system for data processing. The soft X-ray detection unit consists of a parabolic mirror with a proportional counter with a small window area at its focus; for the harder radiation there is a proportional counter with a large window area, which has a mechanical collimator in front of it. The principal technical data are listed in *Table II*.

Before we go on to give a more or less detailed description of the instrument, we shall briefly discuss certain characteristics concerning the sensitivity, the required stability of certain parameters (in particular the gas amplification of the detectors), the minimizing of the background and the determination of the optimum field of view.

Table II. Principal technical data for the Utrecht X-ray measuring instrument.

Total weight	12.6 kg
Electrical power	3 watts
Measuring range	7 channels, 2-70 Å (1.5-2.3; 2.3-3.5; 3.5-4.9; 4.9-7.7; 7.7-20; 20-44; 44-70)
Angle between optical axes of X-ray instruments	3′ (max.)
Field of view of parabolic detector configuration	circular-symetric field, total half-width 40′ or 50′
Field of view of large-area detector	along z-axis: total half-width 50′ and 100′ along y-axis: total half-width 35′
Geometrical area of 44-70 Å detector	140 cm²
Geometrical area of 1.5-44 Å detector	150 cm²

Since the sources we want to measure are in general extremely weak, it is necessary to make the radiation-collecting surface as large as possible. For the long-wave radiation this can be achieved with a fairly small detector by using a grazing-incidence optical system. When X-radiation is incident on a mirror at an angle smaller than the critical grazing angle, total reflection occurs. The critical grazing angle Θ_g is a function of the wavelength: $\Theta_g = C\lambda\sqrt{Z\varrho/A}$, where Z, A and ϱ are respectively the nuclear charge number, the atomic number and the density of the reflecting material; λ is the wavelength of the incident radiation and C is a constant. If the mirror has the form of a paraboloid, the entrance window of the detector can be small. This is important for minimizing the background radiation and increasing the life of the detector: the larger the window the more gas leaks away through it.

The optical system is not efficient for the detection of fairly hard X-radiation; since the critical angle is then very small, the paraboloid would have to be very long in proportion to its width and the geometric surface would still be relatively small. In this energy region a large-area detector is used in combination with a mechanical collimator. The proportional counters used for measuring X-radiation are fitted with extremely thin windows. For soft X-radiation (wavelength 44 Å and longer) polypropylene or a similar material is often used. A feature common to all such materials is that the carbon absorption edge results in a transmission characteristic like that illustrated in *fig. 14*. The value of the transmission at 44.7 Å and the extent of its exponential decrease at longer wavelengths depend on the thickness of the material.

[8] A description of the attitude-control system of the satellite has been given by P. van Otterloo in Philips tech. Rev. 33, 162, 1973.

[9] A description of the onboard computer has been given by G. J. A. Arink in Philips tech. Rev. 34, 1, 1974.

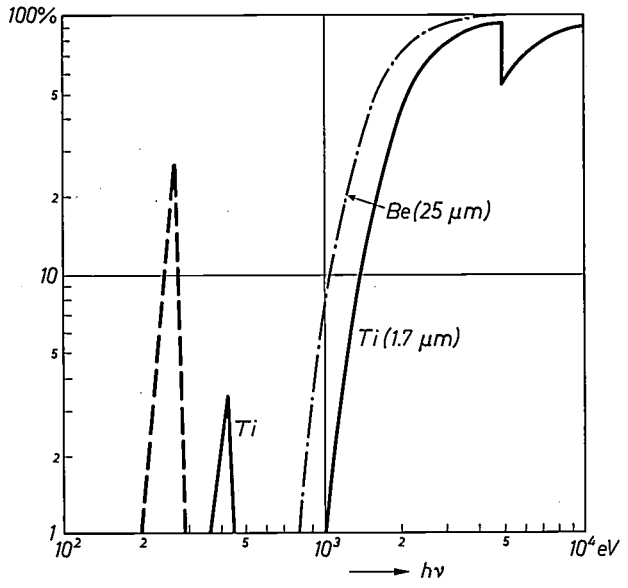


Fig. 14. The dashed line gives the relative transmission of a 3.6-μm thick polypropylene window. Allowance has been made for the effect of the thin layer of aluminium and carbon deposited on it. The values are calculations based on parameters measured on the foil. The solid line gives the transmission of the Ti foil (1.7 μm) used in the detector for the 'hard' range. For comparison the figure also gives the transmission of 25 μm thick Be, the thinnest commercially available beryllium foil (chain-dotted line). This foil is often used in rocket experiments.

The thin plastic windows are not completely gas-tight. The counter gas lost during the flight must of course be replenished, because the gas amplification of a proportional counter is a function of the density of the gas. The gas amplification must remain constant, otherwise the relation between the pulse height at the output of the detector and the energy of the incoming detected photon will gradually change.

Detectors with thin plastic windows have only been employed previously in rockets. The proportional counters used in rockets are of the flow-counter type; these have a controlled leakage, which is made much greater than the leakage through the window. This technique cannot easily be used for satellite experiments because of the large reserve of gas it would require.

The simplest method of controlling the gas amplification was assumed to be to provide the detector with an accurate and reliable pressure meter, and to use its indication for regulating the exact amount of gas to be injected. Since a pressure regulator of this type was not commercially available, we devised the following solution. Part of the detector system was designed in the form of a monitor counter, which is permanently irradiated by a radioactive source. The height of the output pulses from this counter is used as a signal to control the high-voltage supply. Since the gas amplification is also a function of the high voltage on the anode, we compensate the change in gas pressure by a change in the supply voltage [10]. In this way a constant gas amplification can be obtained in a fairly wide pressure range. When so much gas has leaked away that the automatic control can no longer compensate for the gas loss, a command can be sent to the satellite for a quantity of gas to be supplied from the reservoir.

An incidental advantage of this high-voltage control is that the monitor will also react to very intense interfering radiation from the Van Allen belts. When the satellite passes through the radiation belts of the Earth, especially above the southern part of the Atlantic Ocean, where the density of the high-energy electrons is substantially greater than elsewhere at the same altitude, the high voltage will decrease. This will help to prolong the life of the detector. Of course, during this period no X-ray observations can be made.

A problem in the detection of X-radiation is that the detectors are also sensitive to high-energy particles. Much attention has therefore been devoted to minimizing the background radiation. The particle discrimination is based on the difference between the ionization tracks produced by high-energy particles and X-ray photons in a gas. A widely used method is therefore to surround the actual detector with a shield of anti-coincidence counters.

The measurements of cosmic X-radiation previously carried out fall into two categories: on one hand, rocket measurements in the long-wavelength band, usually above 44 Å, and on the other hand rocket and satellite measurements in the short-wavelength band, using detectors that were almost invariably equipped with a beryllium window.

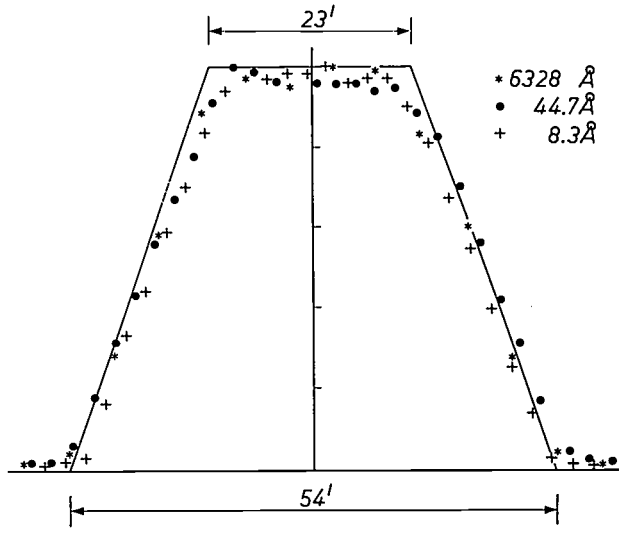


Fig. 15. Angular sensitivity of the collimator for sections *1* and *3* of the large-area detector (see fig. 16) in one direction (in the *x,y*-plane, i.e. the plane perpendicular to the line connecting the Sun and the Earth). The figure gives the theoretical profile (solid line) and the measured profile for visible light and two X-ray wavelengths. The transmission of the collimator together with the grid supporting the Ti window of the detector is 30%.

The transmission curve of a foil of beryllium 25 μm thick has already been shown in fig. 14. The curve shows directly why so few measurements have been carried out so far in the band between 12 Å and 44 Å (1 and 0.22 keV). However, this intermediate band is especially interesting because of the increase of interstellar absorption towards longer wavelengths (fig. 3). Also shown in fig. 14 is the transmission of a 1.7 μm titanium foil. As can be seen, the Ti shows extra transmission between 27 Å and about 35 Å. For this reason the large-area detector (surface area about 12 by 13 cm) is equipped with a Ti window.

Most of the previous experiments, mainly performed with rockets, were designed not only to measure the shape of the spectrum but also, and more particularly, to determine the position of the source. What is then required is a transmission curve in which the sensitivity varies rapidly as a function of angle, e.g. a triangular sensitivity curve.

When the rocket or satellite rotates about its axis, the instrument scans the sky; in this way the position of the source can be accurately determined from the increase and decrease of the radiation intensity. In our case, where the position is assumed to be known and the main objective is to investigate the physical parameters of the sources, an angular-sensitivity curve with a flat top is required; ideally it should be rectangular in shape (*fig. 15*). For various reasons, e.g. if the optical axis of the instrument and that of the attitude-control system do not coincide, it may happen that the desired direction of observation does not entirely coincide with the one actually obtained, and in addition the position of many of the sources to be observed is not exactly known. If the flat top of the angular-sensitivity curve of the instrument is higher than the sum of these uncertainties, the measured intensities can then safely be compared with one another, and we do not run the risk of spurious variations in the observed intensity.

The two detection units

For soft radiation

The detection unit for soft X-radiation consists of the paraboloid mentioned above, with a diaphragm-filter disc in the focus, and close behind it the detector with gas-filling system and reservoir; see *fig. 16*.

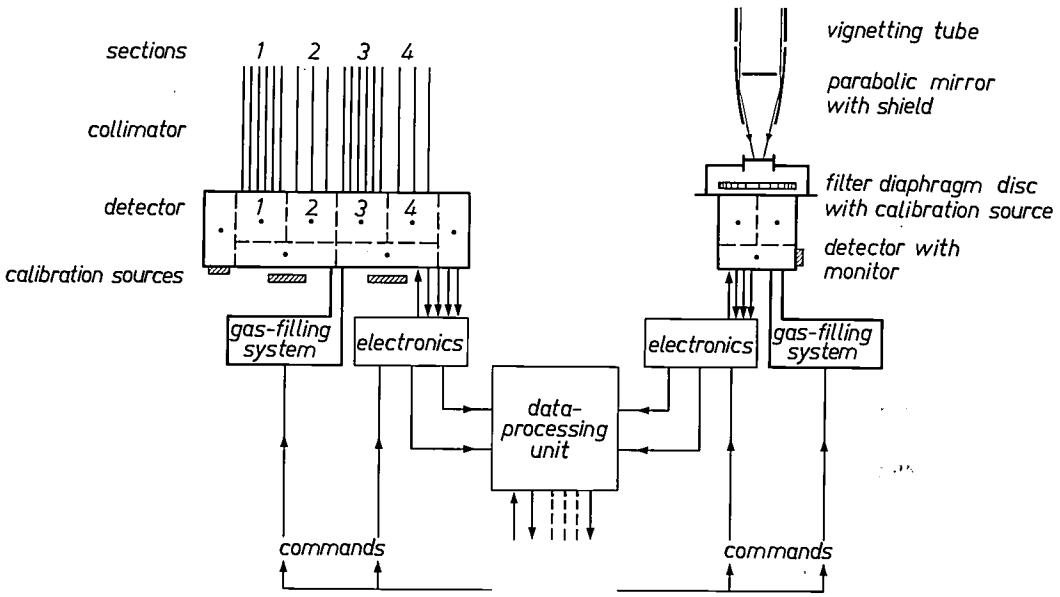
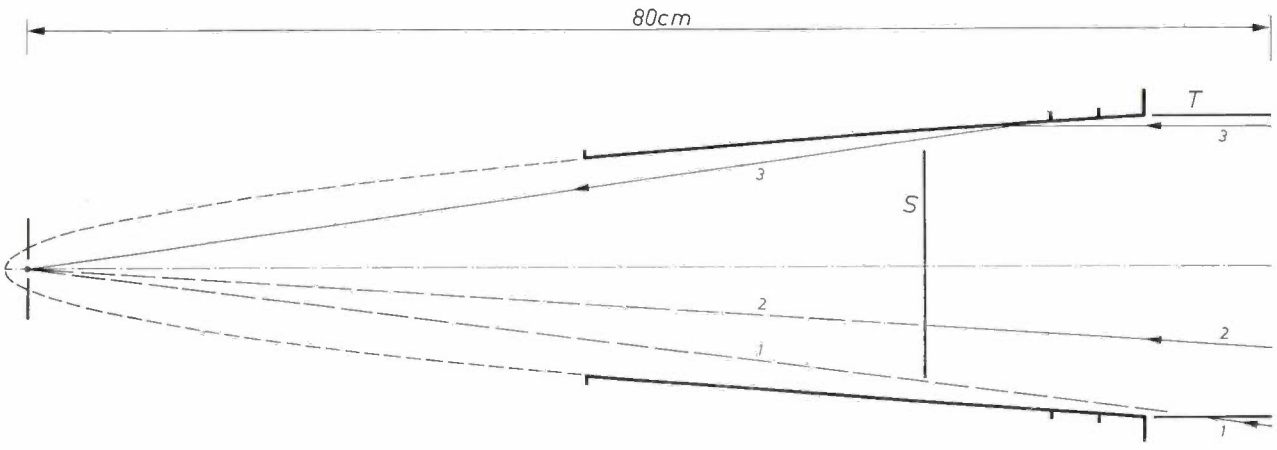


Fig. 16. Diagram of the Utrecht X-ray instrument. *Right*, the detector unit for soft X-radiation (44-70 Å), *left*, the detector unit for harder X-radiation (1.5-44 Å).

The overriding requirement in the design of the paraboloid was of course to obtain the largest possible sensitive area within the limitations of volume and weight. For paraxial rays the angle of incidence is greater the closer they lie to the optical axis. Since reflection only occurs when this angle is smaller than the critical grazing angle, the central part of the paraboloid serves no purpose and can be omitted. The actual paraboloid used is suitable for 45 Å and longer wavelengths. The geometric surface area is about 140 cm². At the centre there is a shield that prevents radiation from reaching the detector directly. In front of the paraboloid is a 6-cm long vignetting tube, whose function is to block out radiation which would otherwise reach the detector

[10] A. C. Brinkman and P. de Groene, Nucl. Instr. Meth. **66**, 316, 1968.



a

b

Fig. 17. *a*) The parabolic reflector for radiation with a wavelength greater than 45 Å. *S* screen. *T* vignetting tube. The paths of three rays are shown. If the tube *T* were not there, ray *1* originating from the cosmic background, or from some source or other, would reach the detector without being reflected from the mirror. Ray *2*, which would reach the mirror without being reflected, is blocked by the screen. Ray *3* belongs to a beam originating from the region of the sky within the desired field of view. The parabolic reflector directs this beam on to the detector. The left-hand part of the paraboloid does not contribute to the reflection. The angle of incidence here is greater than the critical angle.
b) Photograph of the paraboloid. The reflector was designed by the Space Research Laboratory at Utrecht and made by Philips Research Laboratories in Eindhoven. Typical of the dimensional accuracy is the size of the focus for paraxial radiation. The overall half-width of this focus is 0.18 mm. The focal length is 80 cm. The geometric surface is 140 cm^2. The weight is 3.3 kg.

along the shield, without being reflected from the mirror; see *fig. 17*. A diaphragm is situated at the position of the focus. Combination of the diaphragm, the shield and the tube produces an angular-sensitivity curve as shown in *fig. 18*. It should be noted here that a paraboloid is not an optically imaging instrument: paraxial rays are imaged at a point, while rays entering at a small angle form a ring in the focal plane [11]. The paraboloid is therefore an X-ray *collector*.

The proportional counter, situated immediately behind the focus, has a polypropylene window 3.6 μm thick. A thin layer of aluminium (thickness about 200 Å) is deposited on the inside of the window to give a high-conductivity surface; deposited on the outside surface is a layer of graphite (thickness about 0.1 μm), whose function is to absorb ultraviolet radiation.

The detector is filled with 80% argon and 20% CO_2, and consists of three sections: two measuring sections and one anti-coincidence section. A radioactive source is permanently attached in front of the anti-coincidence section, so that it also serves as a sensor for the high-voltage control. All the sections are arranged in anti-coincidence with the others to discriminate against high-energy particles.

The basic diagram of the gas-filling system for the detector is shown in *fig. 19*. The pressure in the detector and the high voltage are measured during the flight;

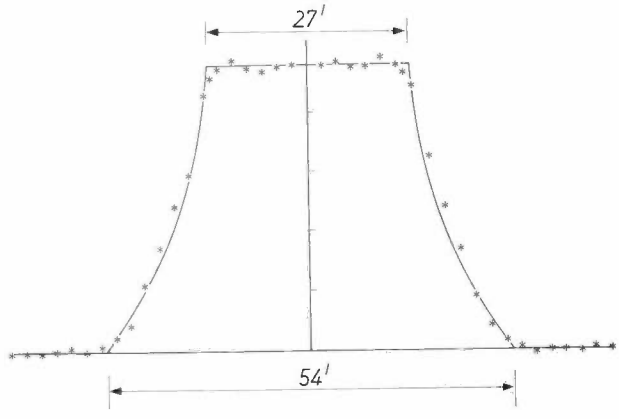


Fig. 18. Angular-sensitivity curve of the paraboloid with vignetting tube, light screen and small diaphragm. The curve again shows the calculated profile; the crosses represent the values measured with visible light.

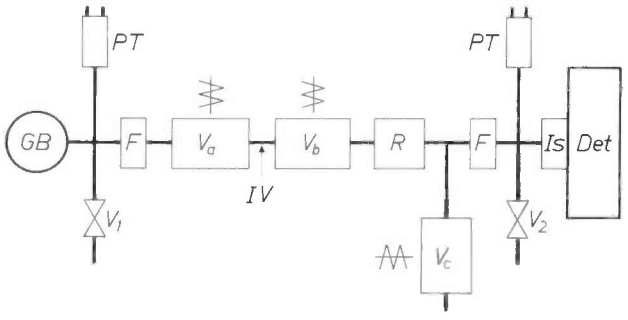


Fig. 19. Principle of the gas-filling system. *GB* gas reservoir. *PT* pressure transducers. *F* particle filters. V_a, V_b, V_c valves. *IV* intermediate volume. *R* restriction. *Is* insulator. *Det* detector. V_1, V_2 manual valves.

every three minutes the measured data is stored in the memory on board the satellite. When the satellite arives above the ground station this data is signalled back to Earth, and in addition these values are measured every 8 seconds during ground contact. If the pressure has dropped too much, the detector gas is replenished. This is done as follows: the valves V_a, V_b and V_c are normally closed, and the intermediate volume *IV* is filled by opening valve V_a. The intermediate volume now contains a quantity of gas at pressure P_1 (75 atmospheres at the launch). One second later the valve V_b is opened. If the pressure in the detector is not yet high enough, the procedure is repeated. As the gas reservoir empties the replenishment has to be more frequently repeated. This is not in itself a problem, since the replenishment and check on the effect can be performed very quickly. Valve V_c serves for completely refilling the detector with fresh gas from time to time.

The diaphragm-filter disc can be set in four fixed positions in front of the detector. In this way it is possible to select two diaphragms for changing the aperture angle, a calibration position and a filter position. In the calibration position the light path to the detector is blocked. This is necessary because during certain periods in flight — after the launching and after a possible eclipse of the Sun by the Earth — a situation could arise in which the Sun was shining into the paraboloid. This would destroy the thin window. In the calibration position the measuring sections are irradiated by a ^{55}Fe radioactive source, which emits almost monochromatic radiation of 2.1 Å. In the fourth position a UV filter is interposed in the light path. In many rocket observations some difficulties have been encountered in the past from UV radiation. The use of a filter (MgF_2) that transmits ultraviolet rays but not X-rays enables the measurement to be carried out twice, with and without the filter, making it possible to distinguish between X-rays and ultraviolet rays.

For harder radiation

The detection unit for measuring the harder radiation, shown in *fig. 20*, consists of a large-area detector fitted with a Ti window. The gas mixture in the detector is Ne-CO_2.

The detector consists of four measuring sections, surrounded by four anti-coincidence counters (fig. 16). To cut down on the number of preamplifiers required, the anodes of section *1* and *3* and the anodes of sections *2* and *4* are directly interconnected. As in the case of the small detector for the long-wavelength range, one of the anti-coincidence sections is in the form of a

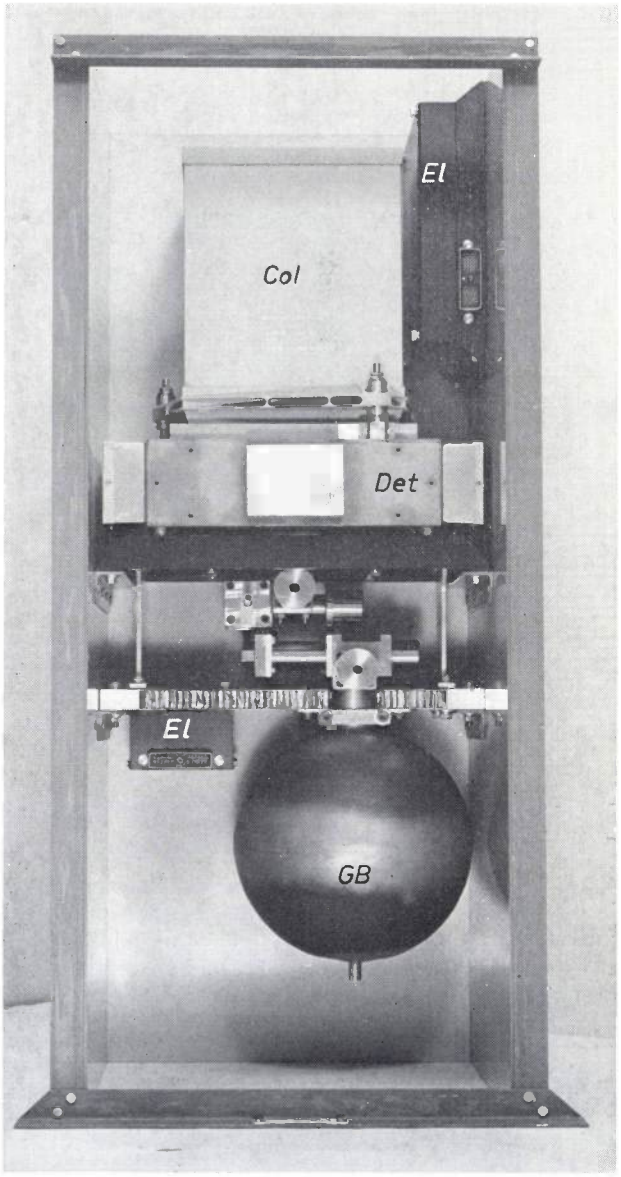


Fig. 20. The large-area array (see fig. 16). *Col* collimator. *Det* detector. *El* electronics. *GB* gas reservoir. Below the detector there is a gas-filling system similar to the one in fig. 19.

[11] R. Giacconi, W. P. Reidy, G. S. Vaiana, L. P. van Speybroeck and T. F. Zehnpfennig, Space Sci. Rev. 9, 3, 1969.

monitor counter, to act as a sensor for the high-voltage control. With metal foils the diffusion through the window is insignificant. It is very difficult, however, to obtain a rolled-metal foil free from pinholes so that here too a gas-filling system of the type just described is used. In the base of the detector there are two 13-μm thick aluminium windows. Two radioactive sources (^{55}Fe and ^{90}Sr) are mounted on a shaft in front of these windows. The shaft is rotated about its axis by a small motor so as to bring the sources in line with the windows. For measurements in the normal mode, i.e. in anti-coincidence between the various sections, only the ^{55}Fe radiation is measured. The gas amplification and the resolution of the detector are directly determined from the position of the peak and its width. On the other hand, when the detector measures events that occur simultaneously in more than one section, only electrons are measured. The electrons from the ^{90}Sr source produce a broad pulse-height distribution, which enables the discriminator levels in the electronics to be determined.

Situated in front of the detector is a mechanical collimator, whose field of view is given in fig. 15. The collimator consists basically of two parts, one for sections *1* and *3* of the proportional counter, and one for sections *2* and *4*. The collimator has a rectangular pattern of slits of two kinds, one measuring 1.4×2.5 mm and the other 1.4×4.85 mm; the bars in between are 0.25 mm thick. This collimator, which is 17 cm long and is composed of thin (1 mm) plates, was designed and made by the Institute of Applied Physics of the TNO-TH organization at Delft. To obtain the flat top in the angular-sensitivity curve, as previously discussed, there is a plate with a different pattern at the centre of the collimator. The two different fields of view make it possible to distinguish between the diffuse X-ray background and the intrinsic background of the detector itself.

Processing of the detector signals; methods of measurement

The pulse height at the output of the detectors is proportional to the energy of the detected photon. The energy spectrum is measured by counting the number of photons and selecting them by pulse height. In our case there are seven pulse-height channels. In deciding on the number of channels we took into account the limited memory capacity of the satellite, the low intensity of the sources and the limited resolution of the detectors. The resolution is inversely proportional to the square root of the energy. The seven channels are: 1.5-2.3 Å; 2.3-3.5 Å; 3.5-4.9 Å; 4.9-7.7 Å; 7.7-20A; 20-44 Å; 44-70 Å. The spectrum of an X-ray source will be determined by fitting a theoretical spectrum to the measured histogram, which will be done by choosing suitable values for the unknown parameters, such as the temperature of the X-ray source. Allowance should of course also be made for the wavelength-dependence of the instrument as a whole.

The instrument has eight memory registers. Seven of these registers, of 12 bits each, are used for storing the number of detected photons per integration period. The eighth register, of 16 bits, is used for storing the command word received from the onboard computer [9]. This determines the operating mode of the instrument, and it is also used for reporting back the result of the command (for example the position of the diaphragm-filter disc). At the end of each integration period, which may be either 1, 4 or 16 seconds, the onboard computer reads the contents of the 12-bit registers, reduces them to eight bits and stores the result in its memory. The reduction process consists in converting from binary to floating-point code (mantissa 4 bits, exponent 3 bits), or in shortening to 8 bits. The required duration of the measurement, the integration period and the conversion in the computer are determined beforehand by the experimenter for each object to be measured.

The attitude-control system has an offset facility for measuring the intensity of the background from time to time during the observation of an X-ray source [8]. The measurement using seven energy channels is referred to as the 'normal mode'. In this mode all pulses registered in the anti-coincidence counters are also stored from time to time.

For calibration the radioactive source is placed in front of the detector. Since the ^{55}Fe source emits near-monochromatic radiation, the resultant pulse-height distribution will be relatively narrow (half-width 18%). For this reason ten separate pulse height channels have been provided to measure this distribution. Since there are seven memory registers in the instrument, one detector calibration consists of two successive measurements in every five channels.

If there should prove to be intense X-ray sources whose spectra show fine structure — and there are indications that this will be the case — similar measurements could be carried out on X-ray sources. Measurements will then be carried out between 1.7 and 3.5 Å in ten channels. We call this the 'high-resolution mode'.

A third operating mode is the 'pulsar mode'. To measure the period of a pulsar we proceed as follows. The time of observation of each individual photon is measured to an accuracy of 1 ms (by counting the pulses from a 1024-Hz clock). Each of the seven registers is used for storing the detection time of one photon. In this way seven photons can be detected in each

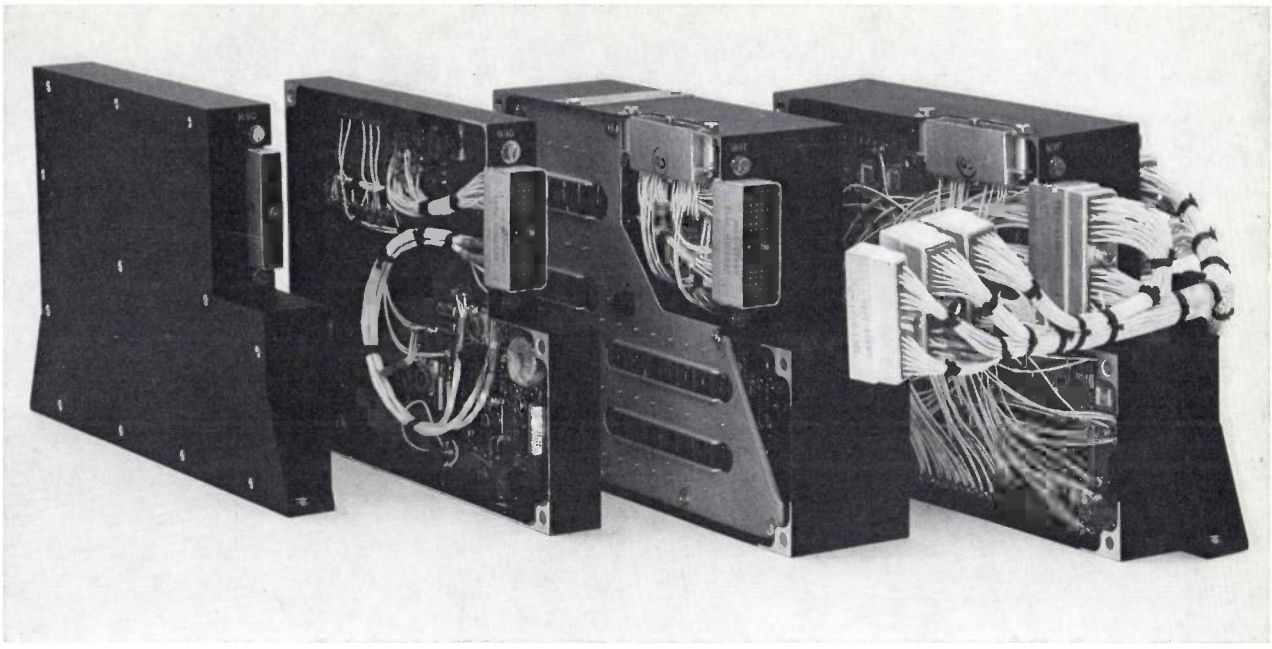


Fig. 21. The pulse-handling circuits of the Utrecht experimental system.

integration period with a minimum of 1 second, but a selection in terms of energy cannot be made at the same time. A rough idea of the spectrum can be obtained by finding out how many photons are counted by one detector and how many by the other.

To determine the time of arrival to an accuracy of 1 ms we need 10 bits. Since the onboard computer is arranged to handle words of 8 bits, it is useful to dispense with the two least-significant bits to make the most efficient use of the memory. We then obtain a resolution of 4 ms, which is sufficient for most cases.

A fourth operating mode, called the 'high-time-resolution mode', was introduced in a late stage of the design. When the results of the first X-ray satellite (UHURU) were known, it turned out that the intensity of many sources was extremely variable, even on time scales smaller than one second. Measurements were therefore needed with a resolving power of better than one second.

With the seven registers we can make seven successive measurements; $\frac{1}{7}$ s in the first register, and so on. In this way we increase the time resolution, though at the expense of the information on photon energies. To simplify the electronic circuits we use periods of $\frac{1}{8}$ s instead of $\frac{1}{7}$ s, which means that we have a dead time of $\frac{1}{8}$ s in every second. Just as in the pulsar mode, we can again select either 2-44 Å radiation or 44-70 Å radiation, or a combination of both. Using the 16-bit command word it is possible to select any required operating mode and any of the four positions of the diaphragm-filter disc by changing the bit configuration. It is also possible to switch off the automatic high-voltage control and to give the high voltage a fixed value: there are two voltage levels to choose from. The preamplifiers can also be independently switched on and off, to permit separate sections of the detector to be used for the measurements.

Fig. 21 shows the electronic units required for the pulse-height discriminators, the decoding of the command word, the memory registers, detector identification, and so on.

Apart from commands using the 16-bit command word, which reach the instrument via the onboard computer, commands can also be given for direct intervention in the operation of the instrument when the satellite is above a ground station.

The American hard X-ray instrument

The hard X-ray instrument on board the ANS satellite was built by American Science and Engineering (AS & E) and the Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, U.S.A. It consists of two instruments: a large-area detector and a Bragg crystal spectrometer. The detector is designed to measure X-ray emission from selected celestial objects and regions in the energy range between 1 and 20 keV, using two closely collimated proportional counters. The Bragg crystal spectrometer will measure two Si emission lines between 1.8 and 2.0 keV, using two Bragg crystals and collimated proportional counters. The instrument is shown in *fig. 22*.

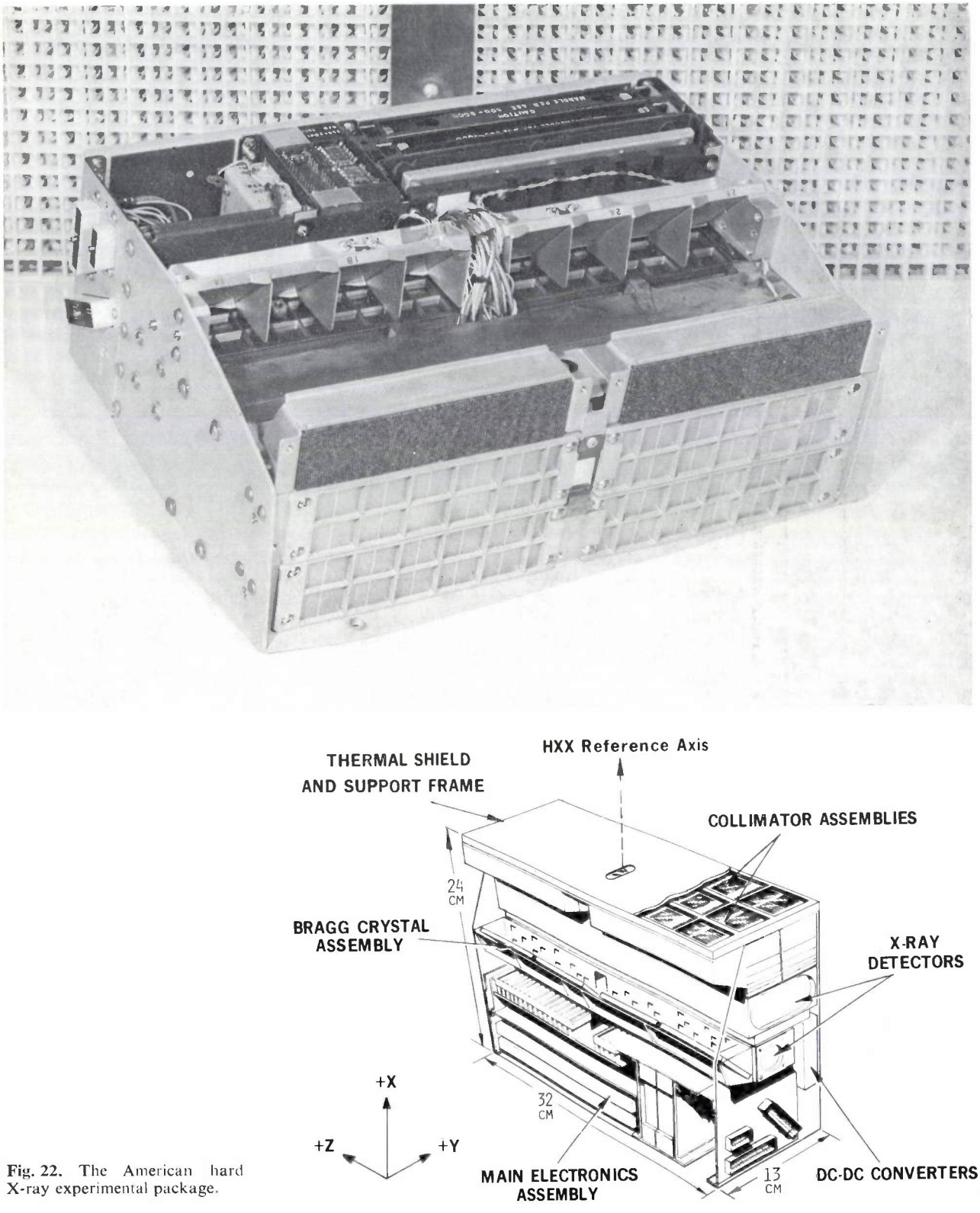


Fig. 22. The American hard X-ray experimental package.

The large-area detector

The large-area array consists of two collimated proportional counters of about 140 cm² each. This area is reduced, however, to an effective 40 cm² each by the losses due to the window structure (76% transmission), the fine collimator (50% transmission) and the coarse collimator (75% transmission). Each detector has a $10' \times 3°$ field of view (half-width), but the optical axes (centre-lines) are offset by 5′, giving an overall field of $15' \times 3°$. The effective total area is thus 60 cm². Differences between the counting rates of the two detectors can yield information on the direction in which

the object is situated, thus providing an azimuthal reference. These differences are found by processing the individual data in the onboard computer. The separate counting rates will also serve as attitude-error signals in the X-ray pointing mode.

It is important that the proportional counters for the large-area detectors should maintain high efficiency over the whole range from 1 to 40 keV. The best filling gas for this purpose proved to be xenon, at a pressure of about 2 atmospheres with a small percentage of quench gas added (9.5% CO_2 + 0.5% He). The gas depth is 3.8 cm, which gives over 90% efficiency between 3 and 16 keV. Each counter has a beryllium window 25 μm thick. The data are analysed in fifteen pulse-height channels. The output of the proportional counters is subjected to pulse-shape discrimination to distinguish between pulses caused by X-rays and pulses due to gamma-ray induced events in the counter. The pulses caused by gamma radiation have a longer rise time than those caused by X-radiation. The aim is to reject about 90% of the gamma-induced pulses while accepting 90% of the X-rays in the 1 to 40 keV range.

The Bragg crystal spectrometer

After extensive investigations it was found that PET crystals — PET meaning $C(CH_2OH)_4$ — are suitable for measuring the line emission of Si, in the energy range between 1.8 and 2.0 keV. There are two crystals, each with an area of about 56 cm^2 and a projected area of about 40 cm^2. They are mounted so as to be sensitive to the spectral lines Si XIII and Si XIV with Bragg angles of 49°50′ and 45°01′ respectively. Independent proportional counters record the reflected X-rays from each crystal. The output pulses are recorded in eight pulse-height channels with the two Si emission-line energies corresponding to the centre energies of two of the eight channels. The relative positions of the two crystals are arranged so that when one of the two crystals is oriented at the critical angle for one of the lines, the other crystal is slightly off the critical angle of the second line. In this way the second crystal receives X-rays corresponding to the X-ray continuum, plus any fluorescence and scattered X-rays from the instrument collimators. Finally, since the large-area detector and one of the two Bragg crystals have the same orientation, it will be possible to measure the photon intensity of the X-ray continuum simultaneously in both detectors.

The Bragg detectors are designed to have a high efficiency near the Si lines. Argon gas at one atmosphere gives an efficiency of 66% for 2.0 keV radiation in a 3-cm thick counter. These counters also have beryllium windows approximately 25 μm thick. The transmission of these windows is about 70% at 2.0 keV and 80% at 2.5 keV. The net detection efficiency will in both cases be about 60%. The background is suppressed both by pulse-shape discrimination and by the use of anti-coincidence techniques. It should be possible to reduce the background counting rate to 0.1 counts per second in each detector.

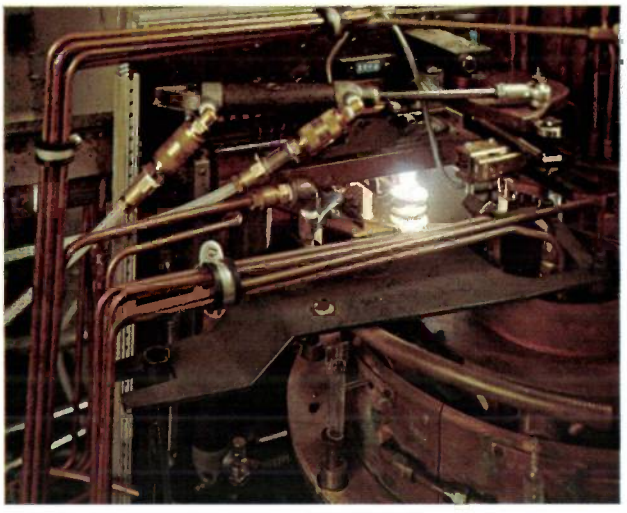
Collimators, calibration and pulse-height analysis

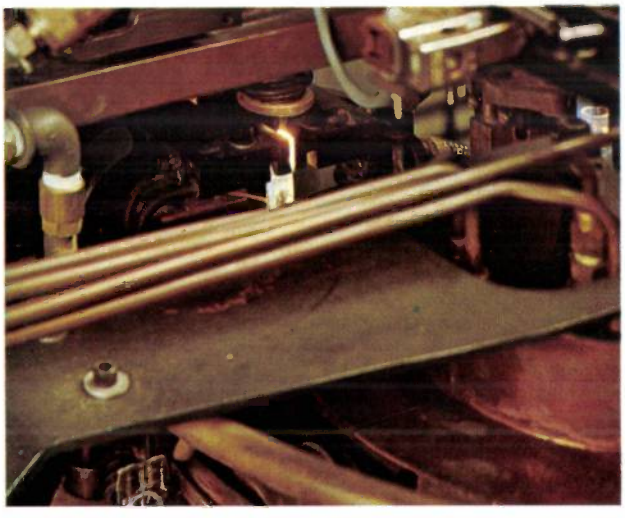
The wire-grid collimators consist of six wire planes mounted in front of each of the large-area detectors. The field of view of each collimator is a slit, with a half-width of 10′. In addition the field of view is restricted by a tube-type collimator of 3° half-width, so that the net field of view is 10′×3°. The tube-type collimator will be used to provide a 3° half-width field of view for the Bragg detector as well. The optical centre-lines of the large-area detector collimator and of the Bragg detector collimator are aligned parallel to within one minute of arc.

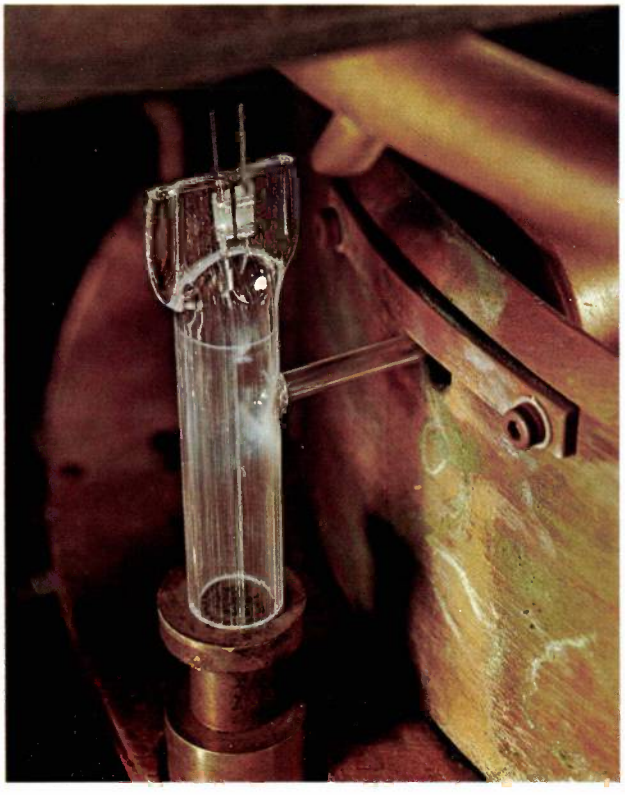
Radioactive calibration sources are mounted in a fixed position in front of each Bragg detector. The sources contain ^{193}Pt, which emits X-rays of around 10 keV. In the calibration mode the gain of the summing amplifier is reduced such that the 10-keV line corresponds to about 4 keV in the normal operating mode of the Bragg detectors.

As mentioned earlier, the pulse-height analyser processes the results of the detector measurements in fifteen energy channels, and those of the two Bragg detectors in eight energy channels. The detector-identification circuit indicates the source of the pulse and delivers control signals to the pulse-height analyser, to permit the data from either the 15-channel system or the 8-channel system to be transmitted and stored at the correct location. In all, the scientific data of this experiment are stored in 27 memory registers of 16 bits each, nominally fifteen for the large-area detector and four for each of two Bragg crystals. The read-out from the memory registers is controlled by the onboard computer.

Summary. The Netherlands astronomical satellite ANS carries on board two experimental packages for measuring the X-ray emission from cosmic sources: one is an Utrecht experiment and the other an American one. The American experimental system, built by American Science and Engineering (AS & E) and the Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, U.S.A., measures X-ray emission in the energy range between 1 and 20 keV. The Utrecht package consists of two instruments, one for the range from 2-44 Å and the other for the range from 44-70 Å. The 'soft' region has not previously been investigated with satellites. Owing to the fairly long pointing periods the ANS will be able to measure photon fluxes with a fair degree of accuracy, even including the emissions from fairly weak sources. A brief account is given of the present state of the art in X-ray astronomy, and of the knowledge acquired of X-ray sources so far measured, many of which — like the pulsars — are variable. The principal features of the instrument are described.







Heating quartz glass with a plasma torch

Extremely high temperatures can be produced with an r.f. argon-plasma torch, whose 'flame' is an ionized cloud of argon obtained by inductive heating. A working temperature of 10 000 °C is easily reached.

Torches of this type are coming into increasing use as emission sources for spectrochemical analysis [1], but they are also very suitable for rapidly heating an object to a high temperature. The photographs illustrate the use of a plasma torch for sealing the electrodes into a quartz-glass tube during the mechanical production of mercury-vapour lamps. The upper photograph shows the heating phase. The tube is near the middle of the picture. The argon gas required for the process is fed in through the pump stem of the quartz-glass tube. The r.f. coil moves downwards until it is around the lower end of the tube. The plasma torch is then ignited, and in about 3 seconds it heats the tube to 1800 °C; the coil then moves up again. Meanwhile, a cartridge in which the electrode is inserted (and perhaps an ignition electrode) is taken from a magazine transported by the slide at the upper left. (The slide can be seen below the hydraulic piston and to its left.) The cartridge moves to the heated end of the tube to place the electrode or electrodes in the correct position, and the tube is then sealed off with pincers (centre photograph). The pincers and cartridge are then moved aside, and the turntable and magazine shift up one position. As can be seen on the far right of both photographs, the tube is preheated with a gas flame immediately before the plasma heating. The bottom photograph shows a finished pinch, into which a discharge electrode and an ignition electrode have been sealed.

Heating with a plasma torch has various advantages over conventional methods. The heating is more homogeneous, giving a stronger pinch. The process is much quicker than with gas burners, and less energy is required; this helps to make conditions in the workshop more comfortable. The plasma torch is also very much quieter than the gas burner. Water is not produced in the heating process, as it is when hydrogen is burnt in conventional burners, so that the quartz glass is thus 'baked' clean, giving a better lamp.

---

[1] P. W. J. M. Boumans, F. J. de Boer and J. W. de Ruiter, A stabilized r.f. argon-plasma torch for emission spectroscopy, Philips tech. Rev. **33**, 50-59, 1973 (No. 2).

Stabilization by oscillation

H. van der Heide

In a corridor of the Philips Research Laboratories in the thirties a most unusual object could be seen: a pendulum balanced upside down on a pivot and held stable in this position by a small vertical oscillation of the pivot. This little-known experiment, due to A. Stephenson, was discovered in the literature by Balthazar van der Pol during his analysis of 'superregeneration', then a new method of radio reception. Equilibrium states analogous to that of the upside-down pendulum are to be found in widely different systems such as a marble rolling in a channel of variable contour, protons in alternating-gradient synchrotons and magnets levitated in combined constant and alternating magnetic fields. The feature common to all these systems is that they can be described by the Mathieu differential equation. The question of how the parameters of such systems should be chosen to obtain stable equilibrium can be answered quite simply with the aid of a stability diagram derived from the Mathieu equation. The use of the diagram will be explained in the article below by means of a few examples. A stable point must be found in the diagram for each degree of freedom. Special attention is paid to levitated magnets. The method described here for the levitation of magnets could also be of interest for frictionless vehicles.

Introduction

According to Earnshaw's theorem it is not possible to achieve a field configuration with permanent magnets such that a magnet can be levitated in stable equilibrium. It may be possible to find a point where the total force and total couple are zero but such an equilibrium is always unstable in one or more directions.

Instability in a given degree of freedom can sometimes be changed to stability by an oscillation. A simple example is an inverted pendulum on a pivot [1]. In *fig. 1*, M represents a mass attached to the end of a rod which is pivoted at the other end A. The situation in fig. 1*a* is stable, that in *b* is unstable. The situation *b* can however be stabilized by setting the point A into forced oscillation in the vertical direction (*c*). A small lateral displacement of M then no longer leads to an increased displacement as in fig. 1*b* but to oscillations as in fig. 1*a*. To achieve this, the forced vertical oscillation of A must satisfy certain strict conditions for amplitude and frequency.

A possibility that now presents itself is that the unstable equilibrium of a magnet levitated in a magnetostatic field could perhaps be made stable in a similar way by means of an alternating magnetic field.

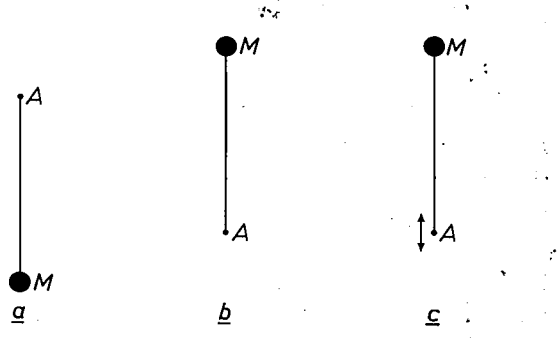


Fig. 1. A mass M attached to one extremity of a rigid rod which is pivoted at the other end A, forms in (*a*) a pendulum in stable equilibrium and in (*b*) a system in unstable equilibrium. The situation (*b*) can be stabilized by oscillation of the pivot A in the vertical direction (*c*).

This type of stability problem may be reduced mathematically to a certain well known differential equation — the Mathieu equation. This equation describes the behaviour of many widely differing physical systems. The solutions of the equation are not simple. However, it is possible, without going into the solutions

H. van der Heide is with Philips Research Laboratories, Eindhoven.

[1] See B. van der Pol, Physica 5, 157, 1925.

themselves, to read off the *stability criteria* from the *stability diagram* that has been derived mathematically from the Mathieu equation. After a brief discussion of the Mathieu equation and the stability diagram, this approach to the stability problem will be illustrated by a number of examples. The last of these examples is the problem mentioned above of the levitated magnet. The situation here is quite complicated because the magnet has six degrees of freedom for each of which a stable point in the diagram has to be found.

The Mathieu equation

Because of gravity the pendulum shown in fig. 1*a* is subject to a restoring force whenever a displacement has taken place. For a small displacement (to which we shall restrict ourselves) the restoring force is proportional to the displacement x: the pendulum has thus a certain lateral *stiffness* C_0 (force per unit lateral displacement) and satisfies the differential equation:

$$m\ddot{x} + C_0 x = 0, \tag{1}$$

where m is the mass of M and $\ddot{x}$ the second derivative of x with respect to time. Equation (1) implies oscillation at the angular frequency $\omega_0 = \sqrt{C_0/m}$. If the pendulum is inverted as in fig. 1*b*, it has a *negative* stiffness equal in absolute value to that of the normal pendulum, fig. 1*a*. As a result the inverted pendulum suffers a displacement that increases as $\exp \omega_0 t$.

The lateral behaviour of the inverted pendulum in case 1*c* is equivalent to the behaviour in the imaginary situation in which A stands still but the acceleration due to gravity has a periodically varying component. If this component varied sinusoidally, the differential equation of the inverted pendulum would be:

$$m\ddot{x} + (C_0 + C_1 \cos 2\omega t)x = 0. \tag{2}$$

(Here ω is not the angular frequency of the vertical oscillation, but *half* that frequency. We shall return to this point of notation later.) Equation (2) is a Mathieu equation. Besides a static stiffness C_0 (negative in the case of fig. 1*c*), the system also has a 'ripple' stiffness $C_1 \cos 2\omega t$. Putting

$$\begin{aligned} \omega t &= \xi, \qquad C_0/m\omega^2 = \omega_0^2/\omega^2 = \alpha, \\ x &= u, \qquad C_1/m\omega^2 = \omega_1^2/\omega^2 = \beta, \end{aligned} \tag{3}$$

equation (2) assumes the canonical form

$$\mathrm{d}^2u/\mathrm{d}\xi^2 + (\alpha + \beta \cos 2\xi)\, u = 0. \tag{4}$$

This quite innocent-looking differential equation has no simple solutions except in the trivial case of $\beta = 0$. In our discussion of this equation we shall assume always that the variable ξ and the coefficients α and β are real.

Strictly speaking the problems to be discussed here cannot usually be reduced to equation (4) itself but only to the more general equation due to Hill:

$$\mathrm{d}^2u/\mathrm{d}\xi^2 + F(\xi)u = 0, \tag{5}$$

where $F(\xi)$ is a periodic function of ξ, of period π, but not necessarily one of the form used in (4). However we shall restrict ourselves below to equation (4) to give a more concrete approach. Also, (4) forms a reasonable model for many problems and many conclusions from it are also valid, at least qualitatively, for (5). If in (5), F(ξ) is a 'square' waveform, we have Meissner's equation, also frequently used as a model.

The stability diagram

In this article we shall be concerned not with the actual solutions of equation (4) but with the question as to whether the solutions relate to stable physical situations. This question is answered by the stability diagram of *fig. 2*. The combinations of α and β in the grey regions of fig. 2 lead to stable solutions; the white regions correspond to unstable solutions [2].

Let us now establish what we mean by 'stable'. According to Floquet's theorem, for all real values of α and β (4) has a solution of the form:

$$u(\xi) = \mathrm{e}^{\mu\xi}\Phi(\xi), \tag{6}$$

where μ is a constant and $\Phi(\xi)$ is a periodic function of period π. Once such a solution has been found, a second independent solution: $\mathrm{e}^{-\mu\xi}\Phi(-\xi)$ is in general immediately obtained, and hence the general solution:

$$u(\xi) = A\mathrm{e}^{\mu\xi}\Phi(\xi) + B\mathrm{e}^{-\mu\xi}\Phi(-\xi). \tag{7}$$

Now in the grey regions of fig. 2, μ is *purely imaginary*. Each solution there is thus a sum of products of bounded periodic functions and is thus bounded over the whole interval $-\infty < \xi < +\infty$. In the regions of instability, on the other hand, μ has a non-vanishing real part, so that there are solutions that 'explode' as ξ goes to $\pm\infty$. On the boundaries the two terms of (7) are not independent; there the general solution is the sum of a bounded and an exploding solution [3].

The inverted pendulum

Let us first conclude the discussion of the stability problem of fig. 1*c*. For the pendulum of fig. 1*a* we have in equation (1): $C_0 = mg/l$, $\omega_0 = \sqrt{g/l}$ (g = acceleration due to gravity, l = length of pendulum); in fig. 1*b*, $C_0 = -mg/l$. If now the pivot A in fig. 1*c* moves vertically as $z = z_0 \cos 2\omega t$, the vertical acceleration is $\ddot{z} = -4\omega^2 z_0 \cos 2\omega t$. The pendulum therefore ex-

periences a vertical force corresponding to an acceleration $g + 4\omega^2 z_0 \cos 2\omega t$ instead of g and therefore a (lateral) stiffness of $-m/l$ times this quantity. In equation (2) we therefore have

$$C_0 = -mg/l, \qquad C_1 = -4m\omega^2 z_0/l,$$

so that (see 3):

$$\alpha = -g/\omega^2 l = -\omega_0^2/\omega^2, \qquad \beta = -4z_0/l.$$

Parametric excitation

To shed some light on the usual factor 2 in the arguments $2\omega t$ and 2ξ in (2) and (4) let us consider the inverse problem for a moment: can a vertical oscillation of the pivot change the stable situation of fig. 1*a* into an unstable one? From fig. 2 we can see that this is so even for very small (infinitely small) β, if $\alpha = 1$, i.e. if $\omega = \omega_0$ (see equation 2). The pendulum can thus be excited into oscillation by an infinitely small vertical

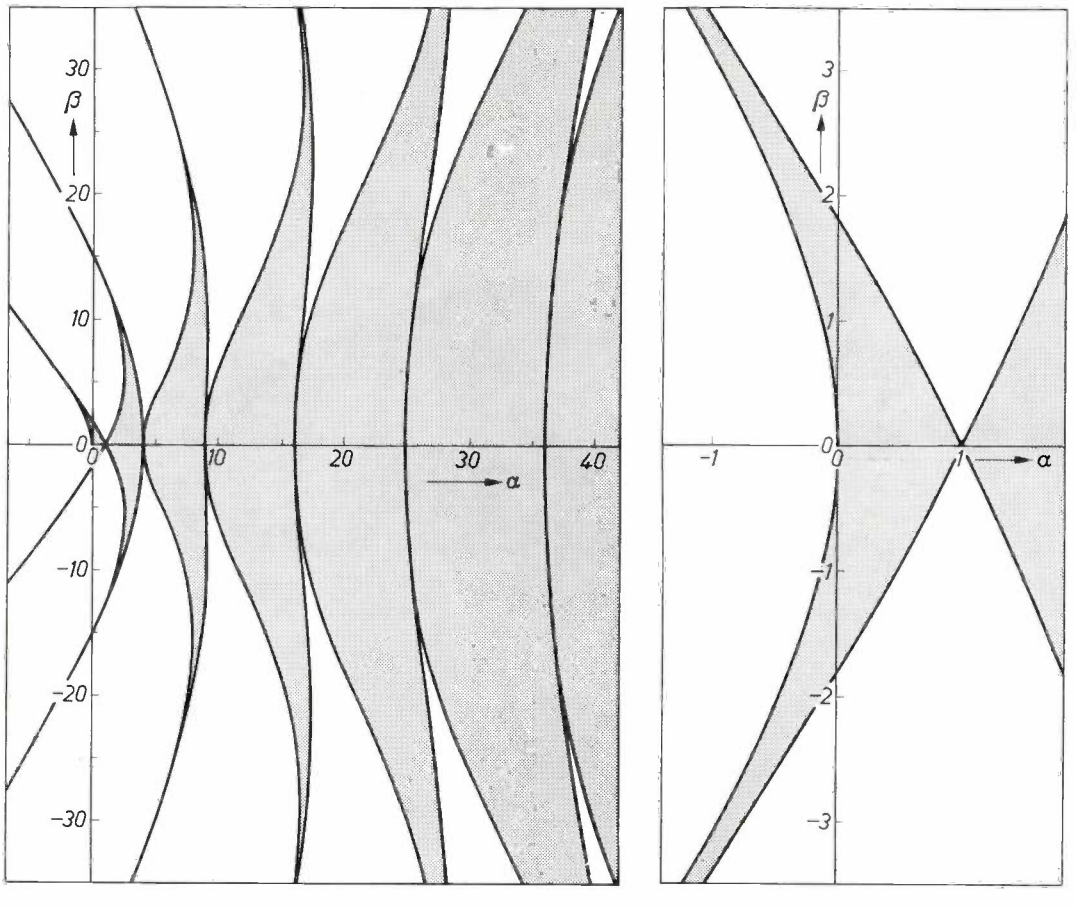


Fig. 2. Stability diagram for the Mathieu equation [2], on two different scales. For the combinations of α and β that lie in the grey regions, the solutions of (4) are stable, i.e. they are bounded for the whole interval $-\infty < \xi < +\infty$. For combinations of α and β that lie outside the grey regions there are instable solutions that 'explode' as ξ goes to $\pm\infty$. The boundaries of the stability regions intersect the α-axis at the squares of the natural numbers.

If we now choose some value for ω, then α is also fixed. We then look at the stability diagram and find a value of β that gives stability for the above value of α, and so find the amplitude z_0 necessary for stability. For example, for $\omega = \omega_0$ ($\alpha = -1$), we find that $|\beta|$ must be about 3.3 to obtain stability (neglecting all stability regions except the first). This implies an amplitude of $\approx 3.3 \times l/4 \approx 0.8\, l$. For a higher frequency ω (smaller $|\alpha|$) the amplitude required is smaller.

[2] The stability diagram shown in fig. 2 and also used in other figures has been calculated with the aid of 'Tables relating to Mathieu functions', National Bureau of Standards, Columbia University Press, New York 1951.

[3] More detailed information on the Mathieu equation can be found in:
J. Meixner and F. W. Schäfke, Mathieusche Funktionen und Sphäroidfunktionen, Springer, Berlin 1954;
N. W. McLachlan, Theory and application of Mathieu functions, Clarendon Press, Oxford 1947;
E. T. Whittaker and G. N. Watson, A course of modern analysis, 4th edition, University Press, Cambridge.
See also the introduction by G. Blanch to the Tables referred to in note [2].

oscillation at a frequency exactly double the natural frequency of the pendulum. This is called parametric excitation [4].

Marble rolling in a channel of variable contour

Let us consider a channel whose cross-sections are parabolae of curvature varying periodically in the y-direction along the channel. Let a marble roll along the channel at a velocity v. In the lateral direction x there is therefore a restoring force of the form $-(A_0 + A_1 \cos ky)x$. The longitudinal distance covered in time t is $y = vt$, so the equation of motion of the marble in the lateral direction is

$$m\ddot{x} + (A_0 + A_1 \cos kvt)x = 0.$$

This Mathieu equation may be put in canonical form (4) by writing:

$$\xi = \tfrac{1}{2}kvt, \qquad \alpha = 4A_0/k^2v^2m, \qquad \beta = 4A_1/k^2v^2m.$$

It follows that for a given channel (A_0, A_1 and k fixed) the ratio β/α is constant and equal to A_1/A_0; for different velocities therefore, a straight line is described through the origin in the stability diagram, *towards* the origin for *increasing* velocity (*fig. 3*). This straight line crosses alternate regions of stability and instability; this is best exemplified by the line *3*. The limiting contours of the six channels corresponding to the six straight lines are shown below the stability diagram. In channel *1* the motion is stable; channels between *1* and *3* give broad stable regions separated by narrow unstable regions; for channels between *3* and *6*, the unstable regions are broad and the stable regions narrow; beyond channel *6* there is no stability. Channels between *5* and *6* correspond to the stability problem of the inverted pendulum, fig. 1*c*; the mean stiffness is less than zero and yet stable regions can be found.

The device shown in *fig. 4* exhibits a close similarity to the marble-in-the-channel problem. It is a simple model to show the existence of the stable and unstable regions. The effective stiffness of the magnet mounted on the leaf-spring varies periodically when the wheel, fitted with alternately magnetized magnets, rotates uniformly. If the wheel is first rotated rapidly (corresponding to a point near the origin of the stability diagram), then, as the rate of rotation diminishes, the magnet is alternately stationary and strongly oscillating.

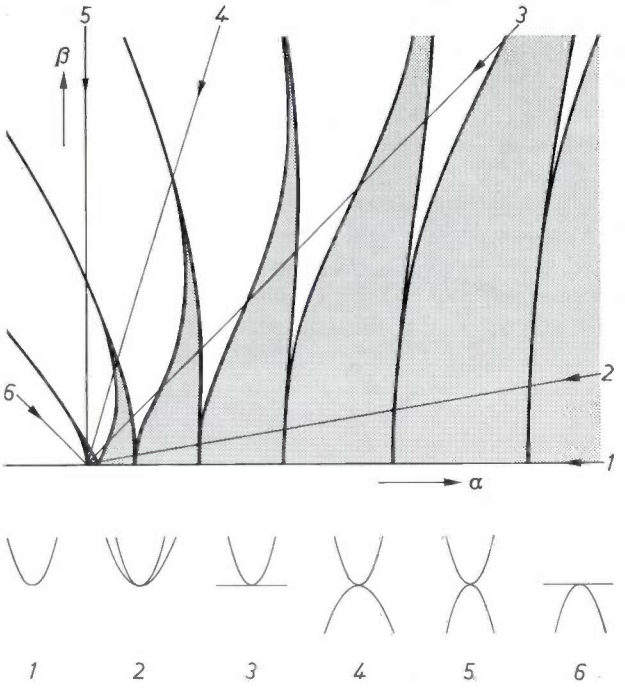


Fig. 3. The lateral stability of a marble rolling with velocity v along a channel of varying profile, depends on the velocity v and on the profiles. A given channel corresponds to a straight line through the origin in the stability diagram; as v increases, this line is traversed towards the origin. In general, then, there are stable-velocity bands alternating with unstable bands. *Below:* Extreme cross-sections between which the profile of the channel varies along its length, for six different channels corresponding to the six lines in the stability diagram. A marble in channel *1* is stable for all velocities; in the following channels the stable regions grow narrower and the unstable regions broader. Beyond channel *6* there are no more stable regions.

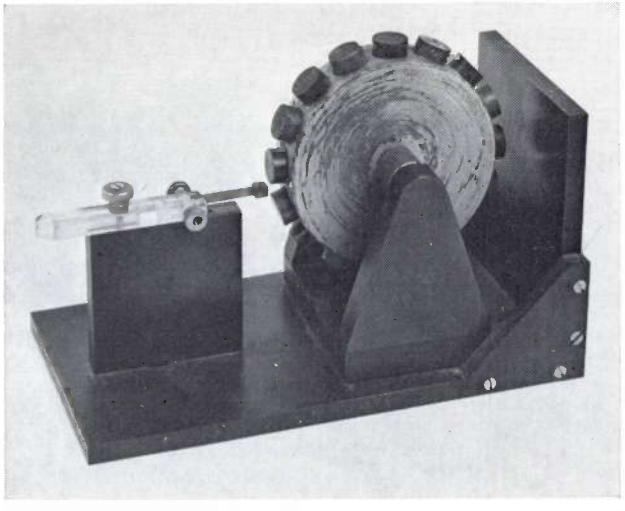


Fig. 4. The magnet mounted on the leaf spring has a periodically varying stiffness when the wheel carrying alternately magnetized magnets is rotated at a uniform speed. In fig. 3 the lines through the origin are traversed away from the origin as the velocity decreases. Passage through an unstable region results in strong oscillations of the magnet; when the velocity corresponds to a stable region the magnet becomes almost stationary. The slope of the straight line traversed depends on the adjustment of the spring and magnet.

Stabilization of the proton orbits in proton synchrotrons (AG focusing)

In the large circular proton accelerators built in the 50s and 60s [5] a problem arises that is closely related to that of the marble in channel number *5* of the pre-

vious problem: the focusing (beam stabilization) of the protons. This is usually done by means of gradients in the 'guiding' field, i.e. the field that guides the particles into the required circular path. An Earnshaw-type situation arises here: a field gradient that focuses in the vertical direction defocuses in the horizontal direction, and vice versa. This may be seen as follows. For simplicity we shall disregard for a moment the guiding field and the curvature of the nominal orbit. The problem is then to hold the proton (of given velocity and momentum) in its (now straight) nominal path; see *fig. 5*. To correct vertical deviations we need focusing Lorentz forces F_v, and hence magnetic fields B_1 (assuming that the beam runs *into* the paper). However, since the B field must be irrotational (curl $B = 0$), components of the type B_1 are always accompanied by components of the type B_2 and these lead to *horizontally defocusing* forces F_h.

A guiding field with a gradient, produced by means of flared pole pieces, may be described as the superposition of a uniform guiding field and a quadrupole field such as that of fig. 5; see *fig. 6*. Here we come to the same Earnshaw-type situation. This can be easily formulated mathematically. The horizontal deflection x and the vertical deflection z under the influence of the magnetic forces are found to satisfy the equations

$$d^2x/ds^2 - (n/R^2)x = 0, \qquad (8a)$$

$$d^2z/ds^2 + (n/R^2)z = 0, \qquad (8b)$$

where s is the position coordinate along the nominal path, R is the distance to the centre of the synchrotron and n is the 'field index' defined by the equality $\partial B/\partial R = -nB/R$. The difference in sign between (8a) and (8b) expresses the difficulty: stability in x implies instability in z, and vice versa.

In the accelerators with weak or constant-gradient (CG) focusing this dilemma is avoided because the centrifugal force has a weak focusing action in the horizontal direction; this was not included in (8a). It can be allowed for by replacing n in (8a) by $(n - 1)$. In this way a positive sign is obtained in both equations if n is taken positive but less than unity. Effectively, there is magnetic focusing in the vertical direction but, by making this weak, the corresponding horizontal defocusing remains smaller than the focusing effect of the centrifugal force.

A revolution in the design of proton accelerators followed the discovery of the possibilities offered by Mathieu's equation (or, rather, by the Hill equation). If, as in CG machines, n is a constant, then (8a) and (8b) can be regarded as degenerate forms of the Mathieu equation, with $\beta = 0$ and $\alpha = \pm n/R^2$. When n is increased from zero, loci are traced in the stability diagram (*fig. 7*) from O to A or B, i.e. *into* the stability

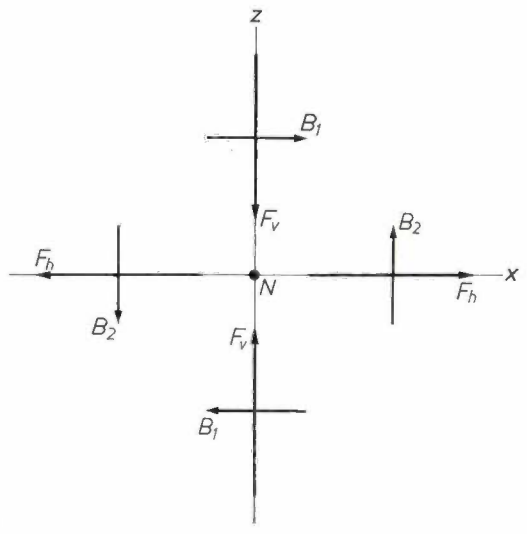


Fig. 5. Earnshaw's theorem for a proton describing a nominally straight path N (perpendicular to the paper). To give it a positive stiffness in the vertical direction (stabilizing forces F_v) by means of the Lorentz force, fields of the type B_1 are necessary but, because curl $B = 0$, these necessarily imply fields of the type B_2 which give a negative stiffness horizontally (defocusing forces F_h).

Fig. 6. The field of flared pole pieces (right) can be thought of as made up of a uniform field (left) — the deflection or 'guiding' field in the case of a synchrotron — and a quadrupole field (centre) as in fig. 5. Since the guiding field contributes no stiffness, the proton is unstable here as in fig. 5.

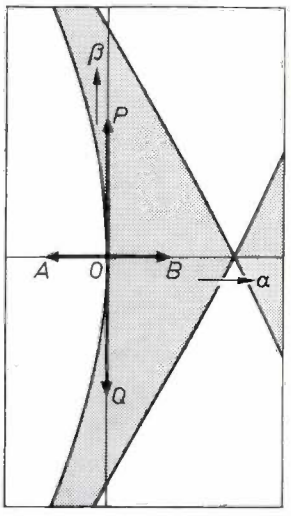


Fig. 7. The nominal path of a proton passing through a series of magnets, all with the same field gradient (constant gradient, CG), is represented by the points A and B in the stability diagram for displacements in the horizontal and vertical directions: in the one direction the equilibrium is stable, in the other unstable. If, however, the magnets have *alternating gradients* (AG), the equilibrium is stable for both directions (points P and Q). This is the basis of the AG focusing of proton beams.

[4] See for example B. Bollée and G. de Vries, Philips tech. Rev. **21**, 47, 1959/60.

[5] See for example R. Gouiran, Philips tech. Rev. **30**, 330, 1969.

region with one equation and *out* of it with the other. When, however, n is not a constant but a sinusoidal function $n(s)$ of s, α is zero and β is not zero; and when the amplitude of $n(s)$ increases from zero loci are traced in fig. 7 from O to P or Q, i.e. into the region of stability for both equations.

In the CERN proton synchrotron at Geneva, whose installation was completed in 1959, and in the Brookhaven alternating-gradient (AG) synchrotron in the U.S.A. completed in 1960, very effective focusing has been achieved on the basis of this AG principle. These machines incorporate a series of guiding magnets with a very large field index (≈ 300), alternately positive and negative. The variation in n is of course not sinusoidal but (partly because of field-free regions in the proton path) much more complicated; however the essential thing is that n varies periodically in sign.

It is rather amusing to see the devious path by which this possibility was discovered in 1952 at the Brookhaven laboratory [6]. Whilst the Cosmotron — the first proton synchrotron (3 GeV) — was nearing completion, studies were already under way for proton machines of higher energies. For deflection and CG focusing in the Cosmotron, C-shaped magnets were used. The useful part of the air gap in these magnets was limited by saturation effects and it was conjectured that the useful part of the gaps could be made larger by placing the yoke of the magnets alternately on the outside and the inside of the ring. The field gradient would then also alternate in sign and its mean would have to have the required value. Calculations were made to check that this gradient variation would not perturb the orbit stability too much but it was found to the surprise of everyone involved that the stability was improved. Further calculations with stronger alternating gradients indicated still better results and this led to the discovery of alternating-gradient focusing (also called strong focusing). The principle had been suggested earlier but had attracted no attention [7].

Electrons in a periodic potential

We shall now consider the quantum-mechanical problem of an electron in an electric field that varies periodically with position. A common example is of course a crystal lattice, in which the periodic field is due to the uniformly spaced ions of the lattice. The charge of all the other electrons may be thought of as smoothed out to a homogeneous continuous charge. This problem is usually attacked not by means of the Mathieu equation but with the Hill or Meissner equations; qualitatively, however, the results are the same. We shall restrict the problem to one dimension. The Schrödinger equation for the wave function $\psi(x)$ of the electron is then

$$d^2\psi/dx^2 + 8\pi^2 m h^{-2}\{E - V(x)\}\psi = 0. \qquad (9)$$

E is the total energy and m the mass of the electron, $V(x)$ is its potential energy as a result of the electric field, and h is Planck's constant. We write:

$$V(x) = V_0 - V_1 \cos(2\pi x/a),$$

where a, the lattice parameter, is the spatial period of the electric field. To reduce (9) to the canonical form (4) we must put:

$$\xi = \pi x/a, \quad \alpha = 8mh^{-2}a^2(E - V_0), \quad \beta = 8mh^{-2}a^2V_1.$$

Since the wave function ψ must be finite for all x, the unstable regions in the α,β-plane now correspond to 'forbidden zones' and the stable regions to 'permitted zones'. We thus see the unfolding of the well known configuration of bands of forbidden and permitted energies: as E increases from $-\infty$ to $+\infty$, we trace out a horizontal line from left to right in the stability diagram (*fig. 8*), whose height β is determined by a and V_1. For small β (small a^2V_1) there is, especially for the higher energies (large α) hardly any separation between the bands of permitted energy; as β increases the energy gaps increase and the bands themselves become narrower.

In the bands of permitted energies (the stable regions), the wave functions of the electron can be written as combinations of functions of the form (6): $e^{\mu\xi}\Phi(\xi)$, where μ is imaginary and $\Phi(\xi)$ has the period π. With $\mu = jka/\pi$ (and k real) and $\xi = \pi x/a$, this expression becomes:

$$\psi(x) = e^{jkx}\Phi(\pi x/a).$$

These are the well known Bloch functions: plane waves e^{jkx}, modulated by a function with the periodicity of the lattice, $\Phi(\pi x/a)$.

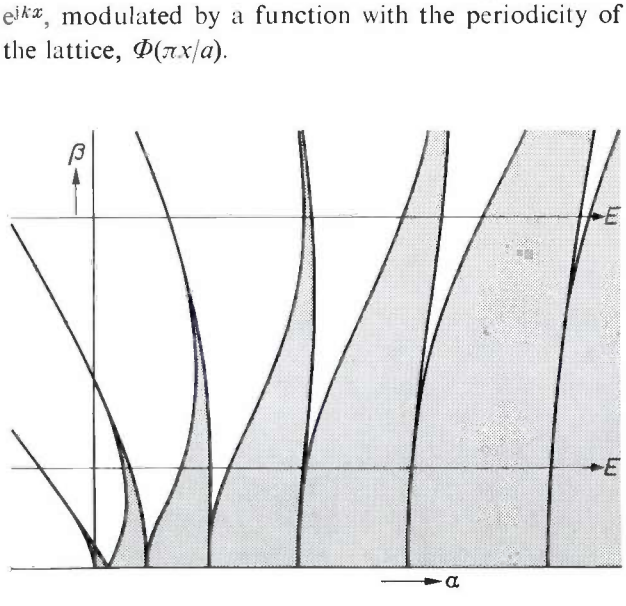


Fig. 8. The Schrödinger wave equation for an electron in a one-dimensional crystal lattice in which it has a potential energy $V_0 - V_1 \cos(2\pi x/a)$ is similar in form to a Mathieu equation. The representative point in the stability diagram moves along a line parallel to the α-axis when the total energy of the electron increases, so that bands of permitted energies (the stable regions) alternate with bands of forbidden energies (the unstable regions). With increasing β, i.e. increasing a^2V_1, the bands of permitted energies become narrower, those of forbidden energies broader.

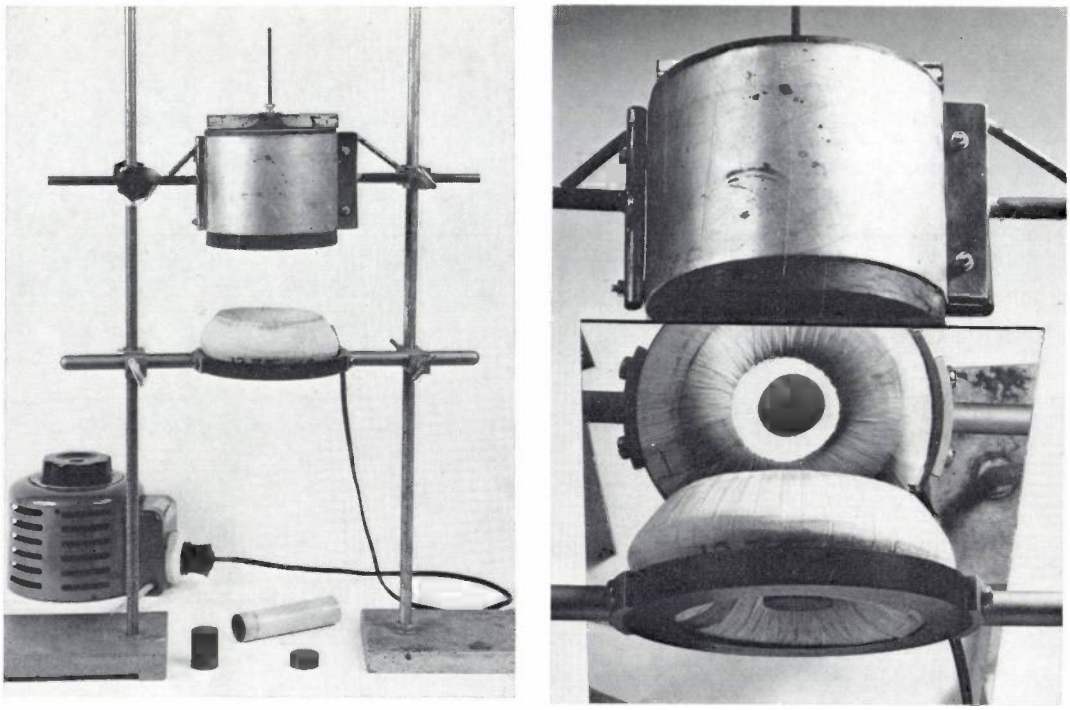


Fig. 9. *Left:* arrangement for the levitation of a magnet. The arrangement consists of a large permanent magnet (above) and a coil carrying an a.c. current. On the table are a small permanent magnet, a stack of three such magnets and a single such magnet provided with a paper 'tail'. The first of these cannot be levitated; the other two can. *Right:* levitation of the triple magnet, made visible by means of a mirror.

Levitated magnets

We shall now give an analysis, by means of the stability diagram, of the equilibrium situation of magnets levitated in the gravitational field by a static and an alternating magnetic field. As an example we shall take the arrangement shown in *fig. 9*. This consists of a permanent magnet and a coil; two of the three magnets lying on the table can be levitated in this arrangement.

In *fig. 10* M_0 represents the permanent magnet of the arrangement and M_1 the levitated magnet. Let us first of all disregard the coil. The origin of the coordinate system x,y,z is taken at the point where M_1, if oriented vertically, is in equilibrium under the influence of gravitation and the attraction of the magnet M_0. The magnet M_1 has six degrees of freedom: three translations A, B and C along the x-, y- and z-axes and three rotations P, Q and R about the x-, y- and z-axes. If M_1 has circular symmetry, the degree of freedom P can be disregarded since this rotation is immaterial and has no effect on the further state of M_1.

According to Earnshaw's theorem the equilibrium state (M_1 at the origin and vertically directed) is unstable. It can be seen directly that the equilibrium is unstable for the translation A: if M_1 moves closer to M_0 the attraction becomes greater so that it flies towards M_0; and it falls if M_1 once moves below the point of equilibrium, $x = 0$. For the translation A there is therefore a negative static stiffness. It is also easy to see that the equilibrium is stable — i.e. has a positive stiffness — for the translations B and C. In fact the sum of the stiffnesses of A, B and C is zero.

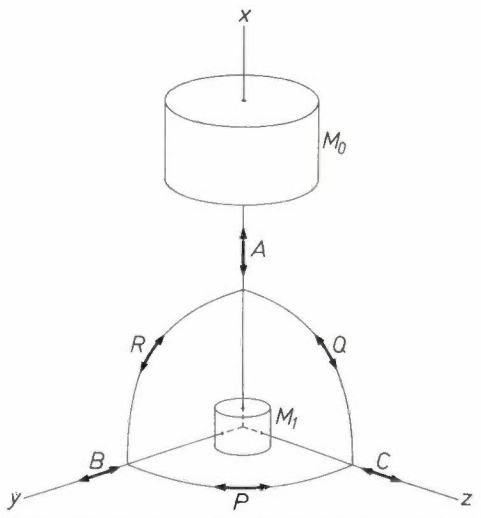


Fig. 10. Coordinate system for the description of the levitation system of fig. 9. M_0 is the fixed permanent magnet, M_1 the levitated magnet in its equilibrium position. M_1 has six degrees of freedom: the translations A, B and C along the x-, y- and z-axes respectively, and the rotations P, Q and R respectively, about these axes.

[6] M. S. Livingston and J. P. Blewett, Particle accelerators, McGraw-Hill, New York 1962, p. 581.

[7] See the article by Gouiran [5], particularly note [5].

This is a general rule for magnets in a magnetic field (and for the gravitational and other inverse-square-law fields): *the sum of the stiffnesses for the three translations is zero* (this is a statement of Earnshaw's theorem).

This sum rule reflects the fact that a particle of mass m and pole strength p subject to a magnetic field and to the gravitational field has a potential energy $w = m\phi_g + p\phi_m$ that satisfies the Laplace equation ($\nabla^2 w = 0$) because the gravitational potential ϕ_g and the magnetic potential ϕ_m (outside the sources of these fields) satisfy the Laplace equation. The potential energy $W = \Sigma(m\phi_g + p\phi_m)$ of a rigid body made up of such particles, as a function of x, y, z, therefore also satisfies the Laplace equation:

$$\nabla^2 W \equiv \frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial x^2} = 0. \qquad (10)$$

This is the sum rule mentioned: the first derivatives of W are (disregarding signs) forces, the second derivatives are therefore stiffnesses.

Earnshaw's theorem is a direct consequence of this rule. It is evident that for a rigid body with only three degrees of freedom (the three translations), no stable equilibrium can be found in free space, since all three stiffnesses have to be positive for stability. This is even more the case for a body with more degrees of freedom, such as a body free to rotate or in which the particles are not rigidly but elastically bound to each other [8]. The theorem is of course also valid for bodies containing charged particles and in an electric field.

The central point of the argument is the fact that a function $W(x,y,z)$ satisfying (10) in a region of space — i.e. a *potential function* — can have no minimum within that space. This has not been proved rigorously in the foregoing; functions can for instance be constructed for which (10) is satisfied at a certain point and which do have a minimum, for example the function $x^4 + y^4 + z^4$ at the point $x = y = z = 0$. The above proposition can however be proved rigorously in potential theory [9].

The magnetic properties of permanent magnets, soft iron, etc., can be simulated for infinitesimal changes by assuming magnetic poles bound rigidly or elastically to the material. The properties of diamagnetic materials ($\mu_r < 1$), on the other hand, cannot be described in this way, for diamagnetism is based on the changes in atomic current loops induced by the magnetic field. Earnshaw's theorem cannot, therefore, be proved for systems containing diamagnetics. Indeed, to the contrary, it can easily be shown that diamagnetic material in a magnetic field of suitable configuration can be levitated in stable equilibrium and this has been confirmed experimentally [10]. Superconductors can be regarded as an extreme case of diamagnetism ($\mu_r = 0$); a spectacular demonstration of this is Arkadiev's experiment — the levitation of a magnet above a superconducting 'dish' [11].

The equilibrium of the levitated magnet M_1 can be made stable for the vertical translation A by means of the coil in fig. 9. With the current in one direction, M_1 is pulled back to the origin (fig. 10) in the event of a vertical displacement; with the current in the opposite direction it is pushed away from the origin. The magnet is thus subjected to an alternately positive and negative stiffness in the x-direction, i.e. to a ripple stiffness when the coil is supplied with a.c. current.

From measurements and calculations a negative *static* stiffness is found for A which, for a distance of ≈ 10 cm between M_0 and M_1 and an a.c. current of 50 Hz, corresponds to a (negative) α_A whose absolute value is much smaller than unity ($< 1/10$). In the stability diagram (*fig. 11*) we see that a ripple stiffness has to be produced that corresponds to a β_A of 1 to $1\frac{1}{2}$, in order to arrive at a point lying centrally in the first stable region. This is possible, although it needs an alternating magnetic field of several thousand ampere-turns. A more fundamental question is that of whether the other degrees of freedom are now perhaps outside the stable regions, since the static and alternating fields produced must give stable equilibrium for *each* degree of freedom.

This can easily be investigated for the translations B and C. Since the sum rule is always valid, it is true for both the static and the ripple stiffnesses. And since B and C are equivalent we find:

$$\alpha_B = \alpha_C = -\tfrac{1}{2}\alpha_A, \qquad \beta_B = \beta_C = -\tfrac{1}{2}\beta_A. \qquad (11)$$

Remembering that the signs of the β's have no significance (as exemplified by the symmetry of the stability diagram about the α-axis), we find the position of B and C as shown in fig. 11. They were already stable in the static field and fig. 11 shows that there is little danger that the alternating field of the coil will upset this stability.

It is another matter with regard to the rotations Q and R. It is clear that in the static field the equilibrium is stable for these (mutually equivalent) rotations. An alternating field that stabilizes A can however easily cause instability in Q and R. Let us look at this in more detail. The static and ripple stiffnesses for A are proportional to $M_1\partial H_0/\partial x$ and $M_1\partial H_1/\partial x$ respectively; for Q they are proportional to M_1H_0 and M_1H_1. H_0 is the field due to M_0 while H_1 is the amplitude of the alternating field. If we now consider a *given configuration* (in which the magnetizations M_1 and M_0 and the current in the coil, i.e. M_1, H_0 and H_1 can still be varied) then $\partial H_0/\partial x$ and $\partial H_1/\partial x$ are proportional to H_0 and H_1. In this case A and Q are very similar, apart from the difference in sign of the static stiffness: for both, the static stiffness is proportional to M_1H_0 and the ripple stiffness proportional to M_1H_1. In reducing the stiffnesses to α's and β's (see eq. 3), however, an important difference arises: where an m (mass) occurs in the translation problem, in the rotation problem we have a J (moment of inertia about the y-axis). In other words, α_A and β_A are proportional to M_1H_0/m and M_1H_1/m respectively, but α_Q and β_Q are proportional to M_1H_0/J and M_1H_1/J. From measurements and calculations it is found that, for an arbitrary magnet M_1,

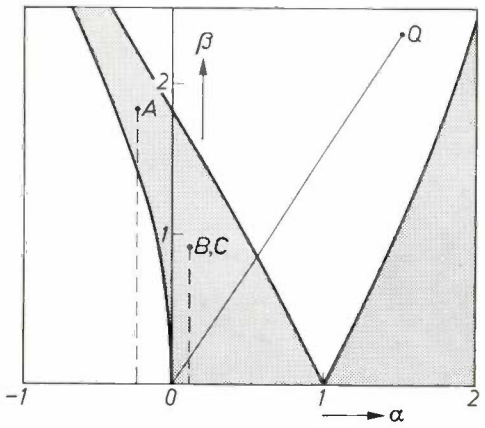


Fig. 11. The equilibrium for the translations A, B and C and the rotation Q represented in the stability diagram for a magnet with a relatively small moment of inertia J in the arrangement of fig. 9. It is assumed that $|\alpha_A| \ll 1$, and that the equilibrium for the translation A is stabilized by means of the alternating field of the coil. The positions of B and C in the diagram then follow from eq. (11); in general they are found to lie in the stable region. The distance of the point Q from the origin is inversely proportional to J. For a relatively small J, Q lies in an unstable region. If J is increased (without changing M_1/m), Q moves towards the origin (while A, B and C remain where they are) so that stable levitation may become possible.

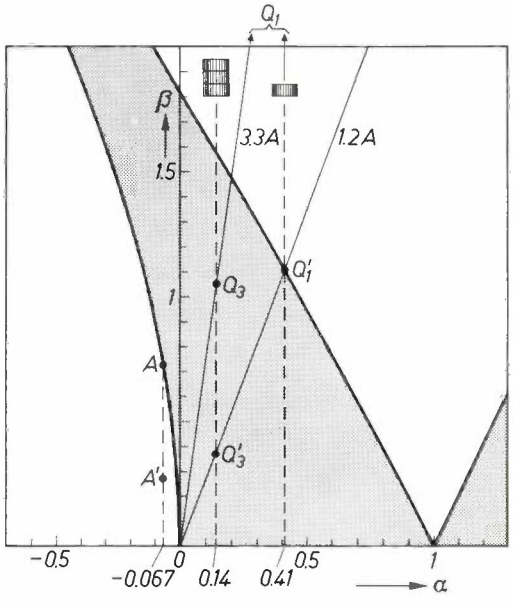


Fig. 12. The equilibrium for vertical translation (A) and rotation (Q) represented in the stability diagram for a single magnet and for a triple magnet, for various currents in the coil of fig. 9. The static stiffnesses $\alpha_A = -0.067$, $\alpha_{Q1} = 0.41$ and $\alpha_{Q3} = 0.14$ were measured with the aid of special mountings and balances that permit movement only in the degrees of freedom A and Q. The values of the ordinate β are all proportional to the alternating field H_1, i.e. to the current in the coil. The locations of Q of the various magnets lie, for a given current, on a straight line through the origin ($\beta_Q/\alpha_Q = H_1/H_0$). At 1.2 A the single magnet becomes unstable for rotation (Q_1'); from this the position of Q_3' follows. Vertical translation becomes stabilized (A) only for a current larger than 3.3 A. For 1.2 A, the vertical translation is represented by A'. At 3.3 A the single magnet lies, for rotation, far in the unstable region (Q_1); the triple magnet, however, still lies in the stable region (Q_3).

J is very easily so small that Q is located in an unstable region if A lies in a stable region (see fig. 11).

Summarizing we can say that a given magnet M_1, in a given configuration and with a given a.c. current, when in stable equilibrium for vertical translation (A, fig. 11) is also in general stable for horizontal translation (B, C) but that it may very well be unstable with respect to rotation about a horizontal axis (Q). If the magnet to be levitated is replaced by another one with the same M_1/m but a smaller M_1/J then A, B and C remain in their same positions but Q moves in the direction of the origin so that the chance of stable levitation increases. Therefore, if a given magnet cannot be stably levitated, then a stack of several such magnets together can be tried: then M_1 and m increase equally, while J increases more rapidly.

This is confirmed by the behaviour of the magnets shown in fig. 9. The single magnet shown cannot be levitated in the arrangement shown whereas three such magnets stacked together can be levitated (fig. 9*b*). An r.m.s. current of at least 3.3 A is necessary for this. The coil has 1250 turns. The properties of the magnets in the arrangement of fig. 9 are shown in *fig. 12*. With the aid of special mountings and balances to limit movements to vertical translation or to rotation about a horizontal axis, we were able to measure the static stiffnesses (see the values of α_A, α_{Q1} and α_{Q3} given in fig. 12) and also to establish whether a given situation would be stable for translation A or rotation Q.

All the β's are proportional to H_1, i.e. to the current in the coil. At 1.2 A the single magnet becomes unstable for rotation; at this current it has therefore reached the point Q_1' and the triple magnet has thus reached the point Q_3'. For vertical translation the magnets are then unstable (A') — the minimum of 3.3 A ($\beta_A = 0.7$) has not yet been reached. At this value of 3.3 A, all β values are 3.3/1.2 × higher than for 1.2 A. For rotation, the single magnet then lies at the point Q_1 and the triple magnet at Q_3. The single magnet is, therefore, very unstable for rotation at 3.3 A whereas the triple magnet is still stable.

In fig. 9*a* another magnet is shown that can be levitated: a single magnet fitted with a paper 'tail'. This tail gives a relative enhancement of the moment of inertia far larger than the relative increase in weight

[8] J. Clerk Maxwell, A treatise on electricity and magnetism, Clarendon Press, Oxford 1873, Vol. I, p. 139.
L. Tonks, Electr. Engng. **59**, 118, 1940.
[9] See for example W. Sternberg, Potentialtheorie, Sammlung Göschen, De Gruyter, Berlin 1925, part I.
[10] W. Braunbek, Z. Physik **112**, 753 and 764, 1939.
A. H. Boerdijk, Philips Res. Repts. **11**, 45, 1956, and Philips tech. Rev. **18**, 125, 1956/57.
[11] V. Arkadiev, Nature **160**, 330, 1947.
See also the title photograph of the article by G. Prast, Philips tech. Rev. **26**, 1, 1965.

so that the position of A in the diagram (fig. 12) is little affected whereas Q_1 is moved considerably in the direction of the origin.

We have devised various combinations of fields in which magnets can be levitated. In *fig. 13* an arrangement is shown in which the alternating field is produced by means of a rotating ring-shaped magnet consisting of four sectors magnetized alternately upwards and downwards. This arrangement requires very little power; the ring can be rotated by a small electric motor permitting the state of stable levitation to continue to exist indefinitely.

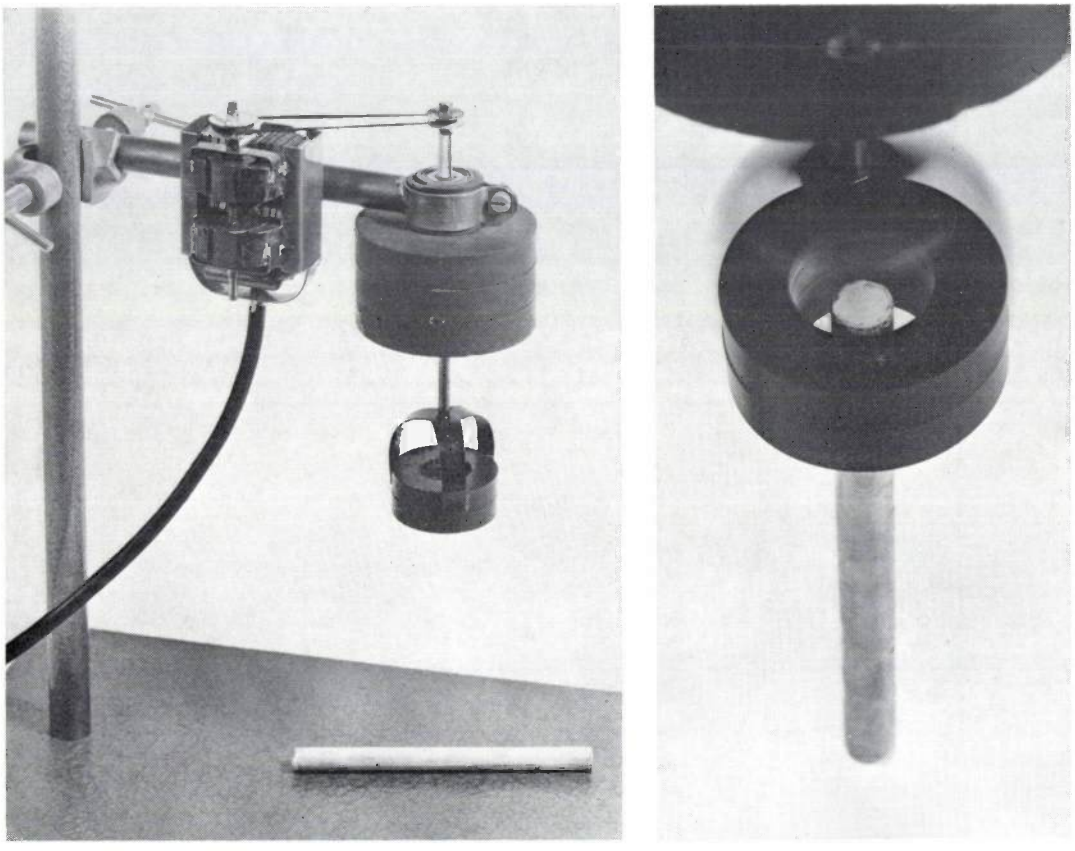


Fig. 13. *Left:* apparatus for levitation consisting of a fixed permanent magnet (the three large ferrite rings) and a rotating ring magnet; the latter is made up of four sectors whose magnetizations are directed alternately upwards and downwards. The magnet to be levitated is provided with a non-magnetic 'tail' to increase its moment of inertia. *Right:* levitation of the magnet.

It was while we were thinking about applications of this kind of magnetic levitation that we designed the model 'train' levitated by permanent magnets as shown in *fig. 14*. The train is supported by the repulsion between magnets in the train and in the rails, M_{0v} and M_{0r} respectively. For vertical displacement there is a positive static stiffness. For lateral movement there is an equally large negative static stiffness; along the rails the stiffness is zero. The lateral instability is eliminated by 'ripple' magnets in the train and in the rails. These magnets give a positive stiffness when the train is at one location (as in the figure) and, one magnet further along the rails, a negative stiffness; when the train is in uniform motion a ripple stiffness is therefore set up. It is clear that with such an arrangement the train can be in stable levitation only when it is actually moving. When starting from rest, therefore, it has to be stabilized laterally by wheels.

It is only fair to say here that levitation by means of permanent magnets — with stabilization by oscillation as in fig. 14 or otherwise — offers no prospects for large systems, i.e. for real trains. The situation is rather like this. Consider an 'infinitely long train' levitated above an 'infinitely long track'; this is then a two-dimensional problem for which the repulsive force per unit length of train can be calculated. This force is found to be proportional to the linear dimension of the cross-section. The weight of the train (still per unit length) is, on the other hand, proportional to the square of the linear dimension. Thus, if the design is

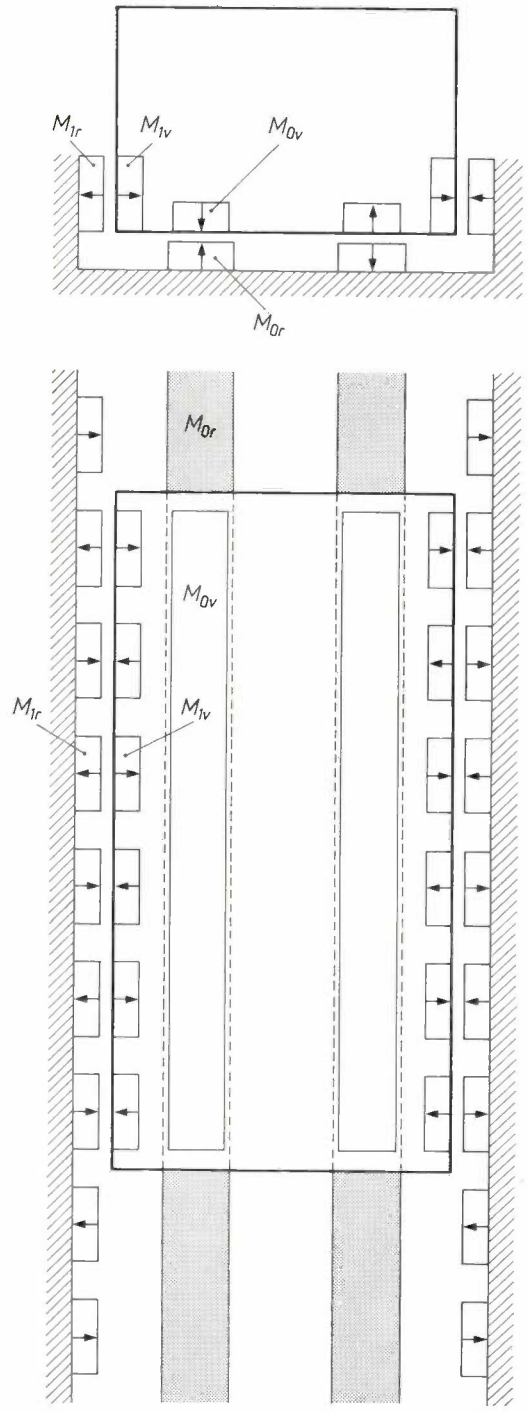


Fig. 14. Schematic configuration for a magnetically levitated train. *Above*, transverse cross-section; *below*, plan. The vehicle is supported by repulsion between the track magnets (M_{0r}) and the vehicle magnets (M_{0v}); it is stabilized in the lateral direction by 'ripple' magnets on the side walls of the track (M_{1r}) and the side magnets in the vehicle (M_{1v}).

scaled up linearly, the weight of the train increases more rapidly than the magnetic force [12]. Equilibrium is therefore achieved only at a relatively smaller separation between train and track; and for large systems this separation is found to be too small to be acceptable.

Damping

Finally a word on the importance of damping in systems described by the Mathieu equation. Suppose that we have a train levitated as in the scheme of fig. 14. The equilibrium for each degree of freedom then corresponds to a stable point in the stability diagram. Possible oscillations about this point (the solutions of the Mathieu equation) can be excited by external perturbations. Such oscillations are undesirable and they must be damped by dissipative elements such as shock absorbers. Besides this function, however, damping has another important advantage: it makes it easier to produce levitation because it broadens the stability regions in the Mathieu diagram. This is evident: with damping the 'weakly increasing' solutions reduce to stable solutions.

This can be shown mathematically [13]. Only an experimental result will be given here, obtained with the apparatus shown in the diagram of *fig. 15*, which is also suitable for measurements of the stability diagram. By means of a d.c. current and an a.c. current through the coils the magnet M is given the required static and ripple stiffness for rotation. Rotation oscillations of the magnet are detected electrically by a small coil. A massive disc S is free to rotate on the shaft. Damping is introduced by means of a viscous oil between disc and shaft. The results of measurements using a particular oil are shown in *fig. 16*.

To illustrate the importance of the broadening of the stability region so achieved, a line has been drawn in fig. 16 (the chain-dotted line through the origin) corresponding to a given ratio of ripple stiffness to (negative) static stiffness. Suppose now that this applies to the lateral movement of the train of fig. 14. If the

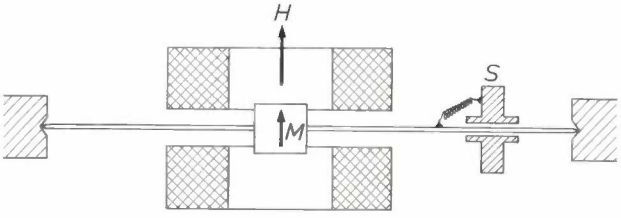


Fig. 15. Apparatus for experimentally determining the stability diagram of the 'Mathieu equation with damping'. A d.c. current and a superimposed alternating current through the coils give the magnet M the required static and ripple stiffness for rotation. Damping of the system is provided by a viscous oil between the disc S and the shaft. The apparatus is intended as a model of the levitated train with damping; M represents a vehicle magnet system and S the rest of the vehicle, attached to each other in the lateral direction (for rotation, in the model) by springs and shock absorbers.

[12] E. R. Laithwaite, Electronics and Power **19**, 310, 1973.
[13] See for example the book by McLachlan [3], p. 96 *et seq.*

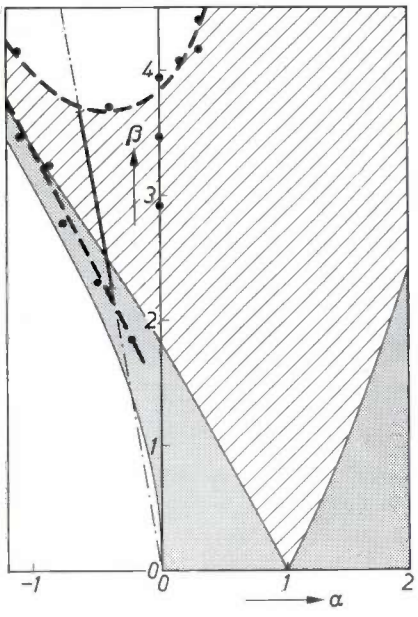


Fig. 16. Stability diagram for the 'damped Mathieu equation' measured by means of the apparatus of fig. 15, using the damping (oil) that gave the most enhancement of the stable regions without complications. Grey: the stable region in the absence of damping (theoretical). Shaded: extension of the stable region as a result of damping (experimental). The bold line represents the region of stable velocities for a train of the type shown in fig. 14 with a given value β/α of the ripple-to-static lateral stiffness ratio, when the train is sufficiently damped in the lateral direction.

velocity increases this line is then traced out towards the origin. It is clear that the range of velocities for which there is lateral stability (the thicker part of the line) is considerably greater than in the absence of damping.

Summary. According to Earnshaw's theorem, the equilibrium of a permanent magnet in the gravitational field and the field of other fixed magnets is always unstable. Unstable equilibria can, however, be stabilized by means of oscillations. This sort of stability problem can be analysed by means of Mathieu's equation. A stability diagram corresponding to this equation can be constructed. Without going into the actual solutions of the Mathieu equation, it is possible to find from the stability diagram the values of the various parameters that give rise to stable equilibrium in a given problem. This is illustrated briefly for the following cases: an inverted 'pendulum', a marble in a channel of periodically varying contour, the focusing of proton beams in synchrotrons and the quantum-mechanical problem of an electron in a crystal lattice. Finally, the equilibrium of magnets levitated in combined static and alternating magnetic fields is analysed with the aid of the stability diagram.

Air-pollution monitors based on chemiluminescence

S. van Heusden

The monitoring and control of atmospheric pollution is a new field, characterized by change and innovation. Until a few years ago it seemed sufficient in most situations to confine the measurements to the SO_2 content in the air, but now it is desirable to know the concentrations of various other pollutants. In the Netherlands, for example, a monitoring network is under construction that will soon enable no fewer than six distinct components — SO_2, NO, NO_2, O_3, CO and H_2S — to be determined.

On the other hand, all the various ways in which these air pollutants can be measured have not yet been fully investigated for their application in continuous automatic monitors. Furthermore, the detectors themselves do not meet the same specifications or requirements in all countries. Very different answers are given, for instance, to the question of how long a monitor should be able to operate without maintenance.

The article below describes four experimental monitors constructed at Philips Research Laboratories in Eindhoven. Designed to detect NO, NO_2 and O_3, their operation is quite different from that of the Philips SO_2 monitor installed in the Dutch national monitoring network. They are all based on chemical reactions involving the emission of radiation. Three of the four detectors do not require the pollutant of interest to be chemically separated from any other pollutants that may be present.

Introduction

In densely populated areas, with heavy concentrations of industry and busy road traffic, it is desirable to monitor the quality of the air by measuring the concentration of at least six pollutants. Apart from the SO_2 concentration, which is already automatically and continuously monitored in various parts of the Netherlands [1], it is important to know how much NO, NO_2 and O_3 the ambient air contains. More attention is also being paid to the concentrations of CO and H_2S [2]. One of the reasons for the importance attached to NO and NO_2 concentrations is that they are a measure of the intensity of motorized traffic, and hence of the other air pollutants it generates, while the main significance of the ozone concentration is that ozone associated with NO and NO_2 is an important factor in the formation of photochemical smog. The concentrations in which ozone, H_2S and the nitrogen oxides occur range from 10 to 500 ppb [*]. The CO concentration is usually much higher, ranging from 1000 to 100 000 ppb.

Instruments capable of measuring the concentrations of these six air pollutants continuously, and without attention, have to meet certain requirements that cannot easily be met by all the methods of analysis that might be used in a laboratory. Indeed, some of these methods have not yet even been investigated to see whether they would be suitable for air analysis, and therefore the pros and cons cannot yet be properly assessed. In any case, the specifications of the detection equipment installed are not everywhere identical [3].

Another method of measurement, besides the coulometric method [1] hitherto used for SO_2 and nitrogen oxides, that looks promising for the measurement of these oxides and of ozone is based on *chemiluminescence*, chemical reactions involving the emission of radiation. A familiar chemiluminescent reaction is the 'glow' of white phosphorus when it is exposed to the air. This effect, which is due to the reaction of the phosphorus with the oxygen in the air, is also observed

S. van Heusden is with Philips Research Laboratories, Eindhoven.

[1] H. J. Brouwer, S. M. de Veer and H. Zeedijk, Philips tech. Rev. 32, 33, 1971.

[2] A nation-wide monitoring network is under construction in the Netherlands that will measure all six components or more if required. See Philips tech. Rev. 33, 194, 1973 (No. 7).

[3] A survey of the priorities given in the Netherlands to the monitoring of pollutants, the methods of analysis that are in principle suitable, and the requirements to be met by automatic monitoring equipment are given in a report: Methods for the automatic measurement of several air pollutants, from the National Institute of Public Health at Bilthoven, the Netherlands.

[*] The abbreviation 'ppb' stands for 'parts per billion', where 'billion' is used in the North-American sense of 10^9, so that 1 ppb = 1 part in 10^9. Similarly, ppm stands for 'parts per million'.

in living organisms, such as the firefly and the glow-worm.

Chemiluminescence occurs when a chemical reaction produces an excited molecule or atom that can give up its energy in the form of a quantum of ultraviolet, visible or infrared radiation:

$$A + B \rightarrow C^* + D \rightarrow C + D + h\nu. \qquad (1)$$

The intensity of the emitted light is proportional to the rate of the reaction, and the spectrum of the light is characteristic of the excited molecule or atom formed, and generally, therefore, of the reaction itself; it seldom happens that two different reactions produce the same luminescent substance.

If such a reaction is to be used for determining the concentration of one of the molecules on the left in equation (1), for example A, then B must be present in considerable excess; the reaction rate is then entirely determined by the concentration of A.

One of the very attractive features of a method of measurement based on chemiluminescence is the ease with which the different air pollutants can be distinguished. Unlike coulometry, in which the detection of these substances is based on a common chemical property — the ability to reduce or oxidize — and requires separation, each substance here is detected through a specific reaction that results in the emission of characteristic radiation. In general, therefore, the measurement of a component is not disturbed by other components present in the air sample; the component to be measured does not have to be separated from the others. It is true that the emitted spectra may overlap, but this difficulty can easily be solved with the aid of optical filters.

Compared with monitors based on the coulometric principle, the advantage of specificity is offset, though not seriously, by the disadvantage of having to consume reagents. In coulometric detectors the active reagent is recovered and the net consumption is zero.

In this article a description will be given of four experimental monitors based on chemiluminescence, which have now been used in our laboratory for more than a year for outside air measurements. There are two detectors that measure the ozone content in the air, one that determines the NO and NO_2 content, and one that measures both ozone and these two oxides of nitrogen.

All these detectors use a photomultiplier tube for measuring the emitted radiation, because of the sensitivity of these tubes in the relevant spectral regions (*fig. 1*).

Each of the detectors complies with the Dutch requirements for the output signal to be proportional to the concentration to be measured over a range of several decades, and for the detection limit to be at about 1 ppb. The measurement time, including the flow times inside the instruments, is no more than a few seconds.

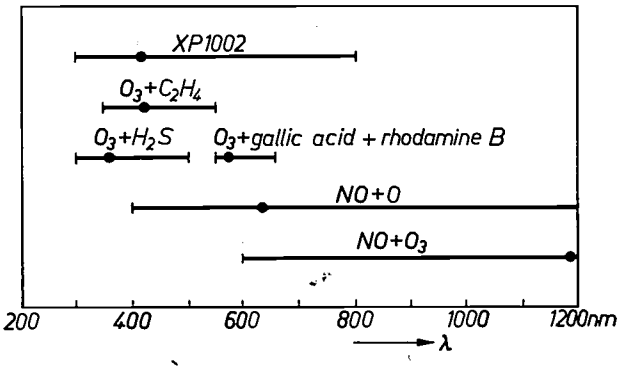


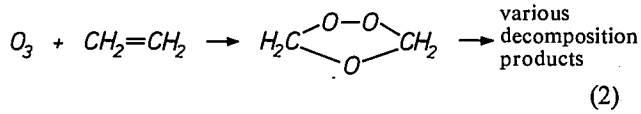
Fig. 1. Situation and spectrum range of a number of chemiluminescent reactions, and the spectral region in which the photomultiplier tube type XP1002 is sensitive. The dots indicate the location of the peaks.

The two ozone detectors already meet the requirement laid down in the Netherlands that a detector should be capable of operating unattended for three months. This is not yet the case with the NO/NO_2 detector, since it has a vacuum pump; at atmospheric pressure the excitation energy would largely be dissipated in the form of heat instead of emitted as radiation. The three-component detector is still in the experimental stage. During the past year it has been found to be capable of working without maintenance for at least two months.

The two ozone detectors

Using ethylene as auxiliary gas

First of all, a detector used for measuring ozone will be discussed, with the aid of *fig. 2*. The auxiliary gas that enters into the luminescent reaction with ozone is ethylene, $CH_2{=}CH_2$ [4]. A pump *P* sucks a constant stream of outside air through the inlet *I*, while ethylene is supplied from the gas cylinder *Cy*. The two flows meet in the reaction vessel *R*, where the following reaction takes place:



(2)

The spectrum of the radiation emitted during this teaction has a maximum at 420 nm (fig. 1). The radiarion strikes the cathode of the photomultiplier tube *Ph*, causing an anode current to flow. No optical filter is used in this detector, because the other common air pollutants, such as NO, NO_2 and SO_2, do not react with ethylene and therefore produce no radiation. The

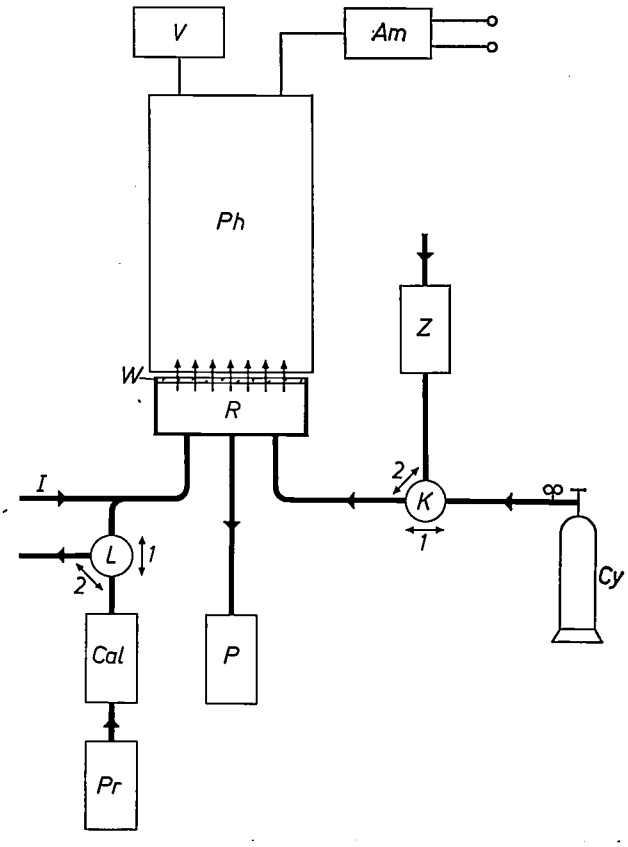


Fig. 2. Diagram of the detector that measures ozone with the aid of ethylene. A pump *P* draws in ambient air through the inlet *I* to the reaction vessel *R*, to which ethylene is also admitted from the gas cylinder *Cy* (valve *K* in position *1*). The light emitted as a result of the reaction between ozone and ethylene is transmitted through a window *W* on to the cathode of the photomultiplier tube *Ph*, causing an anode current to flow which is indicated by the meter *Am* (which can be connected to a pen recorder). The photomultiplier tube receives its power from the voltage source *V*. The valve *K* is set in position *2* at regular intervals to admit outside air from which ethylene and other unsaturated hydrocarbons have been removed by passage through the carbon filter *Z*. With the valve at this setting the 'zero value' of the system is measured, i.e. the dark current of the photomultiplier tube. The system is calibrated from time to time by admitting an exactly known quantity of ozone from the calibration source *Cal* instead of the outside air; this is done by setting valve *L* to position *1* (a stream of air from the pressure cylinder *Pr* prevents admission of the outside air). In position *2* of valve *L* the ozone escapes from the calibration source to the outside air. To ensure reproducibility the calibration source is situated in a thermostat.

various pollutants can however contaminate the window *W* of the reaction vessel. In the ethylene feed line there is a three-way valve *K*, which for ten seconds of each minute is automatically turned to a position in which ethylene is admitted; during the remaining 50 seconds this valve shuts off the ethylene feed and opens to the outside air, which in this case is completely freed from air pollutants by means of a 'zero filter' *Z*. Under these conditions the ethylene is used up so slowly that a gas cylinder of the usual size (50 litres) will last for two years. During the 50 seconds in which purified outside air is admitted, dispelling the residual ethylene and the reaction products, the dark current of the photomultiplier tube is measured. Since the dark current is fairly high compared with the signal to be measured, and varies considerably with temperature, the photomultiplier is housed in a thermostat.

Another important part of the detector is the calibration source *Cal*. Ozone is produced in this source by irradiating a constant stream of clean outside air with a mercury lamp. To ensure that the ozone-forming process remains constant and reproducible, the whole calibration unit is kept at constant temperature. The quantity of ozone delivered varies by no more than 1 % per month. Calibration is effected by supplying a known quantity of ozone to the system for a fixed time by means of the three-way valve *L*.

The detection limit of this detector lies at 2 ppb. The emission intensity is proportional to the ozone concentration up to 1000 ppm (1000×10^{-6}). Regular calibration is necessary since there may be variations in the system parameters on which the output signal depends, such as the flow rate of the monitored outside air, the ethylene flow rate and the supply voltage and sensitivity of the photomultiplier. One calibration every 24 hours is sufficient to limit the error in the measured values to 3 %. The detector is exceptionally fast, and the measuring time can be reduced to 0.1 s. Interference from other components has never been found.

A detector that will measure ethylene and other unsaturated hydrocarbons can also be based on reaction (2); this would require excess ozone.

Using gallic acid and rhodamine B

The second ozone detector has a lower detection limit, but is somewhat more complicated and requires more frequent calibration for it to be sufficiently accurate. The detector is based on the reaction between ozone, gallic acid and rhodamine B [5]. The radiation produced in this reaction has a maximum at 560 nm (fig. 1). On excitation in the reaction with ozone the gallic acid transfers its excitation energy to rhodamine B, which is thus excited in its turn and emits light. The oxidation products of gallic acid are colourless and therefore absorb no light.

The arrangement of the detector can be seen in *fig. 3*. Gallic acid and rhodamine B, deposited on a silica-gel substrate, are contained at *A* in the reaction vessel *R*. The silica-gel substrate has a water-repellent

[4] G. W. Nederbragt, A. van der Horst and J. van Duijn, Nature **206**, 87, 1965.

[5] D. Bersis and E. Vassiliou, Analyst **91**, 499, 1966. This is about an extension of the measuring method based on the reaction between ozone and rhodamine B alone; see V. H. Regener, J. geophys. Res. **69**, 3795, 1964, and R. Guicherit, Z. anal. Chemie **256**, 177, 1971.

layer to minimize the fluctuations in the moisture content caused by varying levels of humidity in the outside air passing over the detector. This is necessary because the quantity of emitted light varies with the moisture content. The outside air is admitted through the inlet *I* by means of a pump *P*. Incorporated in this intake line is a three-way valve *K* which, in position *1*, admits the outside air to be monitored for 8 seconds of every minute. During the remaining 52 seconds, when

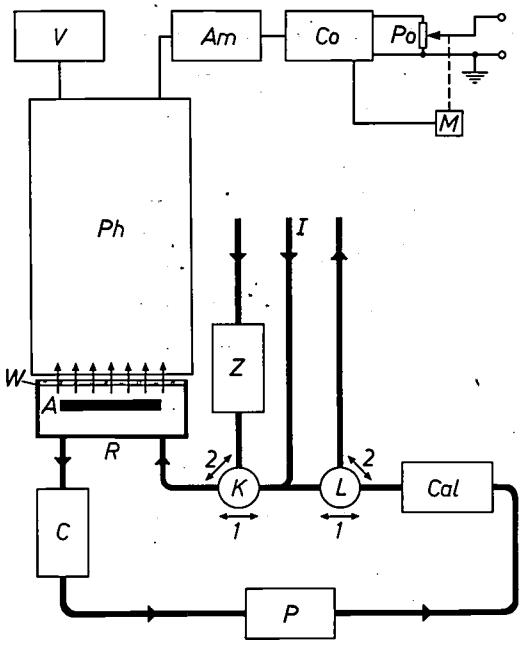


Fig. 3. Diagram of the detector that measures ozone using gallic acid and rhodamine B. *P* pump. *I* inlet. *R* reaction vessel. The two reagents are located at *A* in the reaction vessel, in the form of a deposit on a silica-gel substrate. *Z* and *C* are carbon filters: *C* protects the pump from incoming ozone and allows ozone-free air to be admitted to the calibration source *Cal*. *Ph* photomultiplier tube. *W* window. *V* voltage source. *Am* ammeter. The following measurements are carried out alternately:

valve setting K_1L_2: measurement of ozone in the ambient air,
valve setting K_2 : dark-current measurement,
valve setting K_1L_1: calibration.

The pump action is arranged so that no ambient air is admitted in valve position K_1L_1. For calibration the control circuit *Co*, using a small motor *M*, sets the potentiometer *Po* in a position such that the output signal has a fixed value.

the valve is in position *2*, the outside air is passed through the zero filter *Z* to purify it. During the 52 seconds of the cycle only the dark current of the photomultiplier tube *Ph* is measured. Regular calibration is carried out, for example once every half hour, by means of the three-way valve *L*, using the same calibration source *Cal* as described for the ethylene ozone detector. For the calibration the two three-way valves are in position *1*; the pump action is arranged so that no further outside air is admitted during the calibration, and only the air already present in the lines of the system is circulated. All ozone is removed by a carbon filter *C*, so that ozone-free air enters the calibration source. The calibration is carried out during the period of 8 seconds which is otherwise reserved for a measurement. Once the two three-way valves have been set to the correct positions, 6 seconds elapse before the stream of ozone from the calibration source reaches the reaction vessel, and the actual calibration takes place during the remaining 2 seconds. In this monitor an automatic correction derived from the calibration signal is applied to the measured signal [6]. The output signal is adjusted to the value corresponding to the calibrated concentration of the ozone by a control circuit *Co*, which drives a small motor *M* that controls the setting of a potentiometer *Po*. A particularly important point here is the automatic correction of the change in the reactivity of rhodamine B, which closely depends on the previous history; the reactivity increases with the amount of ozone that has just been measured.

The emission intensity in this monitor is proportional to the ozone concentration up to a value of at least 400 ppb, which is amply sufficient considering that even an ozone concentration of 125 ppb is only a rare occurrence. The detection limit is 0.1 ppb. The life per 'fill' is 500 000 ppb hour. This means that, even with such a high concentration as 100 ppb of ozone day and night (a situation that never occurs), the life of one fill would be 5000 hours, or six months. This particular detector reacts specifically with ozone, the measurement of which is not affected by the normally occurring concentrations of SO_2, NO, NO_2, CH_4, CO and CO_2. *Fig. 4* shows the result of simultaneous recordings of the measurements made by the two ozone detectors (see also fig. 7*a*). As can be seen, the agreement is excellent.

The NO and NO_2 detector

The third detector measures the content of NO and NO_2 in the air. Its operation is based on the following reaction between NO and ozone [7]:

$$NO + O_3 \rightarrow NO_2^* + O_2. \qquad (3)$$

The reaction between ozone and NO_2, in which higher nitrogen oxides are formed, takes place at a negligible rate. For the measurement of the NO_2 content the outside air is passed through special solid reagents, in which the NO_2 is reduced to NO, during part of the measurement time. After this the *total* nitrogen-oxide content (designated NO_x) is determined. The NO_2 content is obtained by subtracting the measured NO content from the content of NO_x.

In the reaction (3) underlying these measurements there is the danger that the excited NO_2 may lose its

energy in the form of heat as a result of collisions; for this reason the reaction is made to take place at reduced pressure. The presence of the vacuum pump used to maintain the low pressure means that this detector cannot operate without attention and maintenance for three months, as required for the Dutch monitoring network.

tions that also occur between ozone and unsaturated hydrocarbons, hydrogen sulphide (fig. 1) or mercaptans. The vacuum pump P gives a vacuum of 270 Pa (2 torr). In the reaction vessel both ozone and NO are admitted. The NO may be obtained by reduction of NO_2, or it can have come directly from the ambient air; the ozone is obtained by means of a dark discharge in

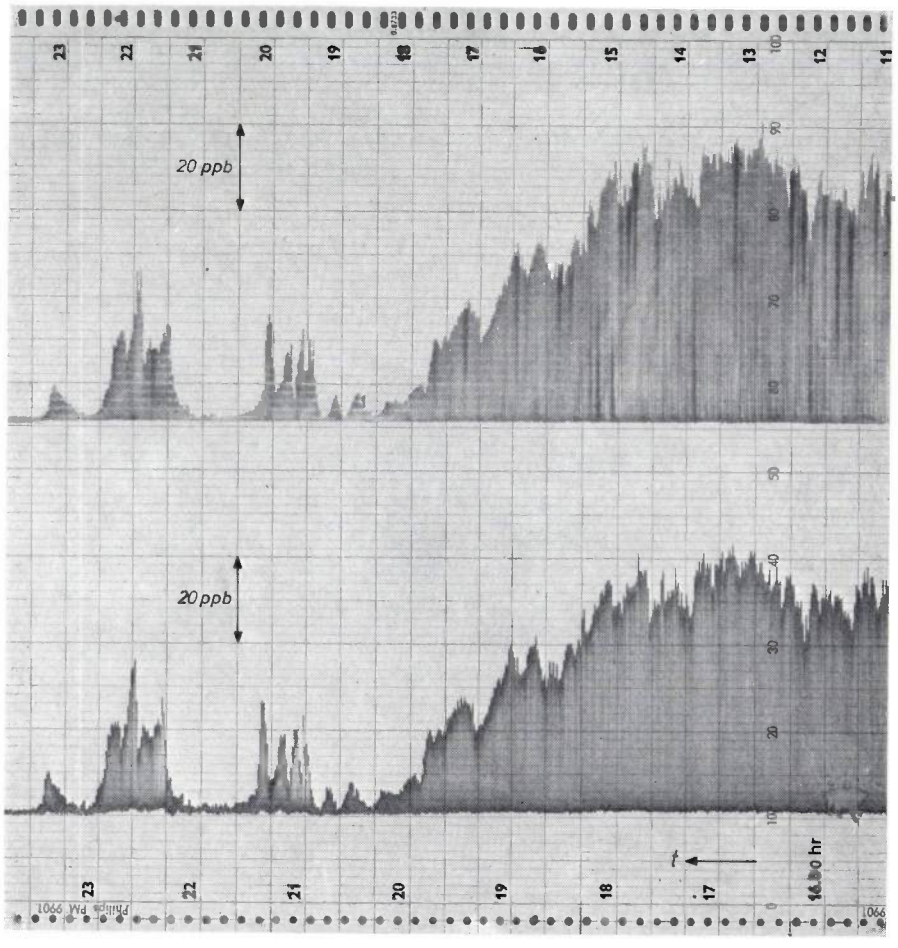


Fig. 4. Recorder traces of ozone measurements with the detector based on the reaction with ethylene (*below*) and the detector based on the reaction with gallic acid and rhodamine B (*above*), recorded in Eindhoven on 9th April 1972 between 15.00 hrs and 24.00 hrs. The traces are virtually identical.

The emitted radiation has a maximum at 1200 nm (fig. 1). The photomultiplier tube used in our detectors is only sensitive to radiation at wavelengths below 800 nm. Although this means that no more than 3 or 4% of the available emission can be measured, it is nevertheless sufficient to achieve a low detection limit.

Fig. 5 shows a diagram of the detector. Between the photomultiplier tube *Ph* and the reaction vessel *R* an optical filter *O* is situated that only passes radiation at wavelengths greater than 600 nm, to avoid possible unwanted effects from radiation at wavelengths below 600 nm. Such radiation can be produced by the reac-

D in a stream of oxygen or air. The 'box' marked *Red* contains the solid reagents that reduce NO_2 to NO. By means of the valves *K* and *L* the following measurements are carried out alternately:

valve setting K_2L_1: NO measurement,
valve setting K_1L_1: dark-current measurement,
valve setting K_2L_2: NO_x (= NO + NO_2) measurement,
valve setting K_1L_2: dark-current measurement.

[6] An automatic correction of this type can in principle be employed in any type of detector.

[7] A. Fontijn, A. J. Sabadell and R. J. Ronco, Anal. Chem. **42**, 575, 1970.

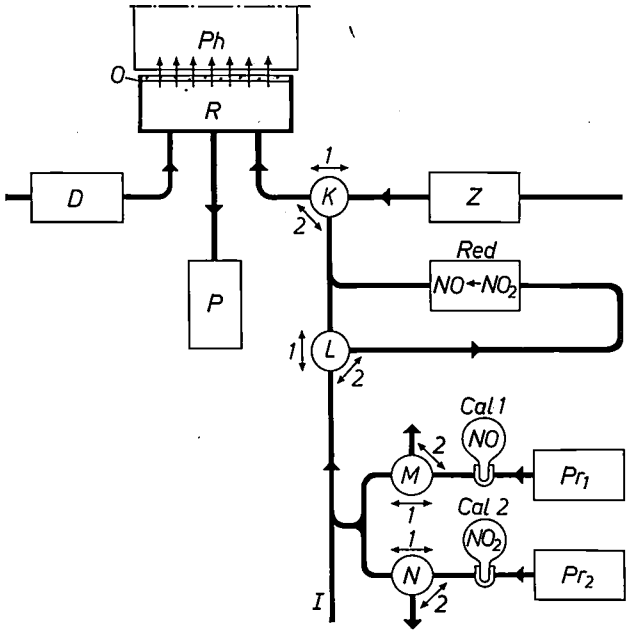


Fig. 5. Diagram of the detector which measures NO and NO_2 with the aid of ozone. The detector is based on the reaction $NO + O_3 \rightarrow NO_2^* + O_2$. Ozone is produced by a dark discharge in a stream of air oxygen in *D*. The NO_2 is measured by first reducing it to NO in the 'box' marked *Red*. With the three-way valves *K* and *L* the following measurements are carried out alternately:

- valve setting K_2L_1: NO measurement,
  valve setting K_1 : dark-current measurement,
  valve setting K_2L_2: NO_x (= NO + NO_2) measurement.

Calibration is effected by means of the three-way valves *M* and *N*, the calibration sources *Cal1* and *Cal2* and the compressed-air cylinders Pr_1 and Pr_2. *O* window of the reaction chamber, also serving as optical filter. The other symbols are as in fig. 3.

Calibration is effected with the aid of valves *M* and *N*. The calibration sources *Cal1* and *Cal2* are permeation sources, consisting of glass spheres that are filled with NO and NO_2 gas respectively, and continuously deliver a constant gas flow through a small orifice covered with permeable silicon rubber. Fig. *7b* shows a recording obtained with this monitor.

The detection limit of the NO/NO_2 detector is 2 ppb. The output signal is proportional to the concentration up to values of 50 ppm. No undesired effects were observed from concentrations of SO_2, CH_4, O_3, CO and CO_2 equivalent to those normally encountered in the outside air.

If NO is supplied in excess, the reaction (3) can also be used for ozone measurements.

The detector for O_3, NO and NO_2

The last of the four detectors measures three air pollutants: ozone, NO and NO_2. Gallic acid and rhodamine B are used in this detector to measure ozone in the way described above.

The NO and NO_2 measurements are carried out by producing ozone from these two substances by a chemical reaction. The NO_2 is decomposed by UV radiation photolysis, and the atomic oxygen produced then reacts with oxygen in the air [8]:

$$NO_2 \xrightarrow{h\nu} NO + O, \qquad (4a)$$

$$O + O_2 + M \rightarrow O_3 + M. \qquad (4b)$$

In the second reaction M is a molecule that takes up the surplus energy, thus preventing the ozone formed from being decomposed in its turn.

The NO is first oxidized with the aid of a solid substance to form NO_2, and ozone is then produced from it as just described.

The photolysis process by which ozone is produced from NO_2 in the detector also occurs in the outside air during the formation of photochemical smog. This type of smog, which irritates the eyes and the mucous membranes, is formed in sunny weather when the concentrations of nitrogen oxides and hydrocarbons are high. This can be the case when there is an inversion layer present in the atmosphere. A characteristic of a smog period is the formation of high concentrations of ozone for reasons that are not immediately apparent. Normally, ozone is primarily formed from oxygen at a great height, in the stratosphere, by the action of ultraviolet solar radiation. The transport of ozone over distances of many miles to lower-lying levels is not very probable, certainly not when there is an inversion layer present in the atmosphere. Another remarkable thing is that the high concentrations of ozone are found more particularly in densely populated urban areas, whereas these are the very areas in which it would be reasonable to expect that the ozone would quickly be destroyed by reactions with other gaseous pollutants and dust. The high ozone concentration comes about as a result of a series of reactions, starting with the reactions (4), the first of which occurs here as a result of solar radiation. The O_3 formed then reacts with NO to form NO_2 by the reaction

$$NO + O_3 \rightarrow NO_2 + O_2, \qquad (3)$$

giving rise to a cyclic process known as the photolysis cycle. This process, which proceeds to an equilibrium:

$$NO_2 + O_2 \overset{h\nu}{\rightleftarrows} NO + O_3,$$

is not in itself sufficient to form high concentrations of ozone. The balance of the equilibrium, however, is shifted to the right by the removal of NO and the addition of NO_2, and this makes high ozone concentrations possible. The shift to the right comes about under the conditions mentioned above in which smog is formed because the NO is then oxidized to NO_2 by various oxidizing substances originating from hydrocarbons. What we have is in fact a second cyclic process: during the photolysis reaction the NO_2 formed gives rise to atomic oxygen, which produces these oxidizing substances as well as ozone. The oxidizing substances, again in combination with NO_2, lead to the formation of peroxyacyl nitrates, which cause the irritation of the eye membranes.

The production of ozone from NO_2 is accompanied by many secondary reactions. The main ones are listed in *Table 1*, together with their reaction rates.

Table I. The photolysis reaction of NO_2 with a number of possible subsidiary reactions [9]. The organic reactions are not taken into account here. The reaction rate R is mentioned for two values of the concentration c of the components NO, NO_2, O_3 and O.

Reaction	R (ppb/min)	
	$c = 10$ ppb	$c = 1$ ppm
$NO_2 \rightarrow NO + O$ (4a)	—	—
$O + O_2 + M \rightarrow O_3 + M$ (4b)	10^5	10^7
$NO + O_3 \rightarrow NO_2 + O_2$ (3)	10^{-3}	10
$NO_2 + O_3 \rightarrow NO_3 + O_2$ (5)	10^{-5}	10^{-1}
$NO_2 + O \rightarrow NO + O_2$ (6)	1	10^4
$NO_2 + O + M \rightarrow NO_3 + M$ (7)	10^{-1}	10^3
$NO + O + M \rightarrow NO_2 + M$ (8)	10^{-1}	10^3
$NO_3 + NO \rightarrow 2NO_2$ (9)	10^{-1}	10^4
$NO_3 + NO_2 \rightarrow N_2O_5$ (10)	10^{-1}	10^3
$2NO + O_2 \rightarrow 2NO_2$ (11)	10^{-7}	10^{-3}

Because of the low rates at which they take place, reactions (5) and (11) are of no significance. The atomic oxygen formed during photolysis (4a) disappears 1000 to 10 000 times faster via reaction (4b) than via reaction (6), (7) or (8). The ozone formed from (4b) reacts a hundred times faster in (3) than in (5), leaving (4a), (4b) and (3) as the principal reactions. The undesirable reaction here is reaction (3). To minimize its influence in the monitor, the location of the photolysis unit is kept close to the gallic acid and the rhodamine B.

A schematic diagram of the three-component detector is shown in *fig. 6*. A fluorescent lamp producing ultraviolet radiation (365 nm) is used for the photolysis, and a helical glass tube around the lamp contains the air mixture which is irradiated for 110 seconds while it remains in the tube (*Rad*). This air mixture should contain NO_2 only, and this is accomplished by passing the mixture through filter F_1 or filter F_2. Filter F_1 breaks down O_3, adsorbs NO_2 and then oxidizes the NO to form NO_2. Filter F_2 breaks down O_3, passes the NO_2 unchanged and oxidizes NO to form NO_2. In the first case the quantity of NO_2 produced is a measure of the NO concentration, in the second it is a measure of the concentration of NO and NO_2 together.

The following measuring modes may be distinguished:

valve setting K_2L_2 : O_3 measurement,
valve setting K_1 : dark-current measurement,
valve setting $K_2L_1M_1$: NO measurement,
valve setting K_1 : dark-current measurement,
valve setting $K_2L_1M_2$: NO_x measurement,
valve setting K_1 : dark-current measurement.

In practice the ozone is measured twice as often as NO or NO_x, to facilitate the reading of the chart recordings.

Regular calibration is required, for example, to correct for variations in the intensity of the photolysis,

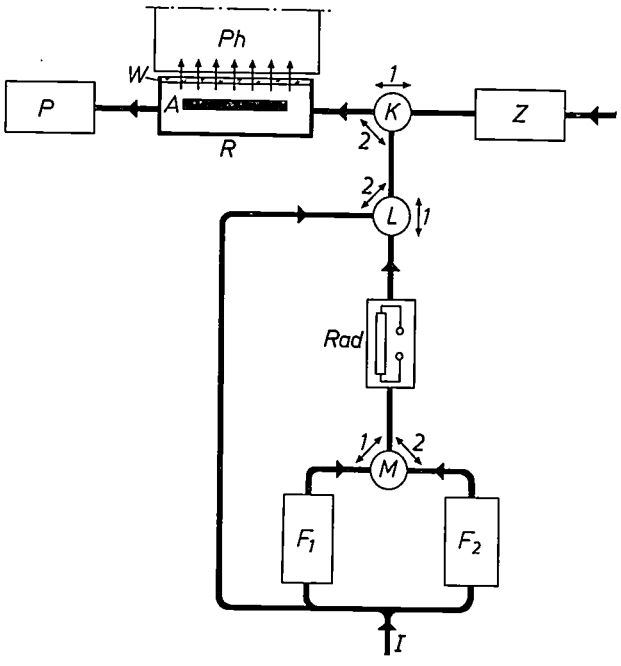


Fig. 6. Diagram of the detector that measures O_3, NO and NO_2 using gallic acid and rhodamine B. The NO_2 is measured after ozone has been produced from it photolytically by irradiation in *Rad*. The chemical filter F_1 breaks down the O_3, adsorbs NO_2 and oxidizes NO to NO_2. The chemical filter F_2 breaks down the O_3, passes NO_2 unchanged and oxidizes NO to NO_2. When the air flows through F_1 the NO concentration is measured, when the air flows through F_2 the NO_x concentration is measured. The following measuring modes may be distinguished:

valve setting K_2L_2 : O_3 measurement,
valve setting K_1 : dark-current measurement,
valve setting $K_2L_1M_1$: NO measurement,
valve setting $K_2L_1M_2$: NO_x measurement.

and in particular because reactions may take place in the equipment between hydrocarbons and atomic oxygen that could give rise to substances capable of oxidizing NO into NO_2.

The complete measurement cycle lasts 4 minutes. The detection limit for ozone is 0.1 ppb and for the nitrogen oxides 0.5 ppb. The output signal is proportional to the measured concentration up to 150 ppb.

Fig. 7 shows three recorder tracks made simultaneously with the three-component monitor, the ozone monitor based on the reaction with gallic acid and rhodamine B, and the monitor for NO and NO_2. As with the two ozone detectors (fig. 4) the agreement is seen to be very good.

It can be seen that the two ozone detectors more than satisfy the severest requirements at present imposed, and are therefore suitable in all respects for practical

[8] See for example: Air quality criteria for photochemical oxidants, U.S. Department of Health, Education and Welfare, 1970, and: R. Guicherit, Atmos. Environ. 6, 807, 1972.
[9] E. A. Schuck and E. R. Stephens, in: Advances in environmental sciences Vol. 1, ed. J. N. Pitts Jr. and R. L. Metcalf, Wiley-Interscience, New York 1969, p. 73.

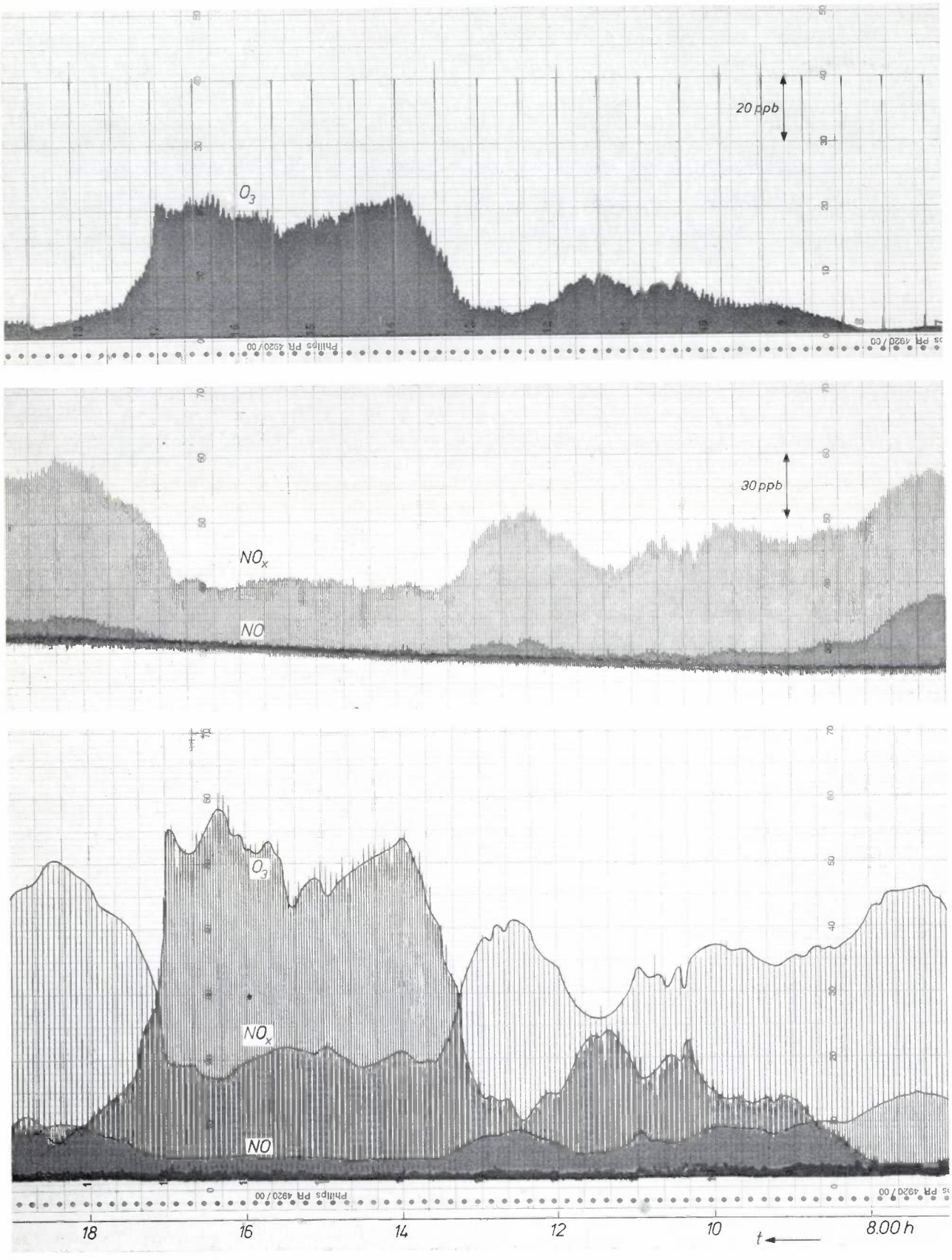


Fig. 7. Recordings made on 29th May 1973 from 07.00 hrs to 19.00 hrs in Eindhoven with the ozone detector based on the reaction with gallic acid and rhodamine B (*a*), with the NO and NO_2 detector (*b*) and with the three-component detector (*c*). In (*a*) a calibration peak of 80 ppb was also recorded once every half hour. Lines are drawn through the concentration values in (*c*) to make the curves more visible.

application. Whether further development of the NO/NO_2 detector will also yield an alternative to the now much improved coulometric devices in situations where the somewhat shorter maintenance interval is acceptable remains as yet undecided.

Summary. Chemiluminescence is the emission of radiant energy from a material as the direct result of a chemical reaction. The emitted radiation is in general characteristic of the one reaction. If therefore the intensity of the radiation is used as a measure of the chemical components taking part in this reaction, a high degree of specificity is guaranteed. The article describes four monitors based on chemiluminescence for measuring air pollutants, in particular ozone (O_3), NO and NO_2. The first monitor determines the ozone content by making the ozone react with ethylene. The light produced is measured with a photomultiplier tube. The second monitor, also for ozone, is based on the reaction between ozone, gallic acid and rhodamine B. The third monitor measures NO and NO_2 and is based on the reaction between NO and O_3. To measure NO_2 as well, the NO_2 is reduced to NO. In the fourth monitor O_3 is measured with the aid of gallic acid and rhodamine B, and NO_2 is made to produce O_3; the NO is first oxidized to form NO_2 and O_3 is then produced from it in a similar way. The detection limit of all four monitors is of the order of 1 part in 10^9, the output signal is proportional to the concentration over several decades, and the total measuring time is no more than a few seconds.

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.

C. S. Aitchison & R. Davies: A varactor-tuned *Q*-band Gunn oscillator.
Int. J. Electronics **35**, 105-108, 1973 (No. 1). *M*

V. Belevitch: On the asymptotic behaviour of meromorphic *RL*-impedances.
Nato Advanced Study Institute on Network and Signal Theory, Conf. Bournemouth 1972, pp. 240-247; 1973. *B*

K. Board: Gunn throws down the gauntlet to IMPATT.
Electronics Weekly, June 20, 1973, p. 20. *M*

A. J. van Bommel & J. E. Crombeen: A LEED and AES study of the PH_3 adsorption on clean Si(111).
Surface Sci. **36**, 773-777, 1973 (No. 2). *E*

C. A. Bosselaar (Philips Semiconductor Development Laboratories, Nijmegen): Charge injection into SiO_2 from reverse-biased junctions.
Solid-State Electronics **16**, 648-651, 1973 (No. 5).

M. R. Boudry: Theoretical origins of N_{ss} peaks observed in Gray-Brown MOS studies.
Appl. Phys. Letters **22**, 530-531, 1973 (No. 10). *M*

P. C. Brandon & O. Elgersma: Effects of α-benzyl-α-bromo-malodinitrile on the primary electron acceptor of Photosystem II in spinach chloroplasts.
Biochim. biophys. Acta **292**, 753-762, 1973 (No. 3). *E*

H. H. Brongersma & P. M. Mul: Absolute configuration assignment of molecules and crystals in discussion.
Chem. Phys. Letters **19**, 217-220, 1973 (No. 2). *E*

R. Brun, J. Cayzac & R. Genève: Reduction of the distortion caused by reflections in mobile radio-relays for colour-television outside broadcasts.
E.B.U. Rev. tech. Part No. 137, 14-22, 1973. *L*

K. H. J. Buschow: Phase relationships and magnetic properties of uranium-gallium compounds.
J. less-common Met. **31**, 165-168, 1973 (No. 1). *E*

K. Carl & K. Geisen: Dielectric and optical properties of a quasi-ferroelectric PLZT ceramic.
Proc. IEEE **61**, 967-974, 1973 (No. 7). *A*

A. L. Dalisa & R. J. Seymour: Convolution scattering model for ferroelectric ceramics and other display media.
Proc. IEEE **61**, 981-991, 1973 (No. 7). *N*

H. Damsma & E. E. Havinga: Influence of a small lattice deformation on the superconductive critical temperature of alloys with the Cu_3Au-type structure.
J. Phys. Chem. Solids **34**, 813-816, 1973 (No. 5). *E*

M. Davio: Taylor expansions of symmetric Boolean functions.
Philips Res. Repts. **28**, 466-474, 1973 (No. 5). *B*

J. P. Deschamps & A. Thayse: On a theory of discrete functions, Part I. The lattice structure of discrete functions.
Philips Res. Repts. **28**, 397-423, 1973 (No. 5). *B*

J. W. F. Dorleijn & W. F. Druyvesteyn: Analysis of the magnetic bubble collapse method.
Appl. Phys. **1**, 167-169, 1973 (No. 3). *E*

E. Dormann (Technische Hochschule, Darmstadt), **K. H. J. Buschow, K. N. R. Taylor** (University of Durham, U.K.), **G. Brown** (Univ. Durham) **& M. A. A. Issa** (Univ. Durham): A study of the lineshape of the NMR spin echo spectra of the compounds $Gd_{1-x}Y_xAl_2$ and $Gd_{1-x}La_xAl_2$.
J. Physics F **3**, 220-232, 1973 (No. 1). *E*

G. Engelsma: Induction of phenylalanine ammonia-lyase by dichlobenil in gherkin seedlings.
Acta bot. neerl. **22**, 49-54, 1973 (No. 1). *E*

C. J. Gerritsma, J. A. Geurst & A. M. J. Spruijt: Magnetic-field-induced motion of disclinations in a twisted nematic layer.
Physics Letters **43A**, 356-358, 1973 (No. 4). *E*

J. J. Goedbloed: Noise in IMPATT-diode oscillators.
Thesis, Eindhoven 1973. (Philips Res. Repts. Suppl. 1973, No. 7.) *E*

H. C. de Graaff: Collector models for bipolar transistors.
Solid-State Electronics **16**, 587-600, 1973 (No. 5). *E*

G. Groh, E. Klotz & H. Weiss: Simple and fast method for the presentation of the two-dimensional modulation transfer function of x-ray systems.
Appl. Optics **12**, 1693-1697, 1973 (No. 7). *H*

P. Guétin & G. Schréder: Phonon tunneling spectroscopy in *n*-Ge Schottky barriers under pressure.
Phys. Rev. B **7**, 3697-3702, 1973 (No. 8). *L*

W. Gutknecht & W. Pies: Verfahren zur automatischen Prüfung von kugelförmigen Kernreaktorbrennelementen.
Materialprüfung **15**, 229-232, 1973 (No. 7). *H*

S. H. Hagen & A. W. C. van Kemenade: Investigation of exciton complexes in 6H-SiC by isotope substitution.
J. Luminescence **6**, 131-136, 1973 (No. 2). *E*

P. Hansen: Magnetostriction of ruthenium-substituted yttrium iron garnet.
Phys. Rev. B **8**, 246-253, 1973 (No. 1). *H*

P. Hansen & J.-P. Krumme: Determination of the local variation of the magnetic properties of liquid-phase epitaxial iron garnet films.
J. appl. Phys. **44**, 2847-2852, 1973 (No. 6). *H*

C. M. Hargreaves: Radiative transfer between closely spaced bodies.
Thesis, Leiden 1973. (Philips Res. Repts. Suppl. 1973, No. 5.) *E*

J. C. M. Henning, J. H. den Boef & G. G. P. van Gorkom: Electron-spin-resonance spectra of nearest-neighbor Cr^{3+} pairs in the spinel $ZnGa_2O_4$.
Phys. Rev. B **7**, 1825-1833, 1973 (No. 5). *E*

L. Heyne: Mixed ionic and electronic conduction in solids.
Fast ion transport in solids, solid state batteries and devices, Proc. Conf. Belgirate 1972, editor W. van Gool, pp. 123-139; 1973. *E*

L. Hollan & A. Mircea: Multi-layer epitaxial structures for high efficiency GaAs IMPATT diodes.
Gallium Arsenide and related compounds, Proc. 4th Int. Symp., Boulder 1972, pp. 217-223; 1973. *L*

K. Hoselitz: Slow growth needs research.
Physics Bull. **24**, 335, 1973 (June). *M*

S. van Houten: Physical aspects of displays.
Trends in Physics, Lectures 2nd General Conf. Eur. Phys. Soc., Wiesbaden 1972, pp. 153-175; 1973. *E*

W. P. A. Joosen: Finite-length effect in a solid-rotor motor.
Philips Res. Repts. **28**, 485-495, 1973 (No. 5). *E*

E. T. Keve & A. D. Annis: Studies of phases, phase transitions and properties of some PLZT ceramics.
Ferroelectrics **5**, 77-89, 1973 (No. 1/2). *M*

M. P. Koster (Philips Centre for Technology, Eindhoven): Vibrations of cam mechanisms and their consequences on the design.
Thesis, Eindhoven 1973. (Philips Res. Repts. Suppl. 1973, No. 6.)

E. Krätzig & W. Schreiber: Ultrasonic investigation of proximity effects in superconductivity under the influence of magnetic fields.
Phys. kondens. Mat. **16**, 95-106, 1973 (No. 2). *H*

U. Krüger, R. Pepperl & U. J. Schmidt: Electrooptic materials for digital light beam deflectors.
Proc. IEEE **61**, 992-1007, 1973 (No. 7). *H*

M. Laguës, J. L. Domange (E.N.S.C.P., Paris) **& J. P. Hurault:** Segregation of dissolved Li to the surface of Si: a new activation process.
Solid State Comm. **12**, 203-206, 1973 (No. 3). *L*

R. E. van der Leest: The coulometric determination of oxygen with the electrolytically generated viologen radical-cation.
J. electroanal. Chem. interf. Electrochem. **43**, 251-255, 1973 (No. 2). *E*

J. W. Orton: Measurement of the geometrical transverse magnetoresistance effect in n-type GaAs at high temperatures.
J. Physics D **6**, 851-859, 1973 (No. 7). *M*

R. F. Pearson: Application of magneto-optic effects in magnetic materials.
Contemp. Phys. **14**, 201-211, 1973 (No. 3). *M*

K. R. Peschmann, J. T. Calow & K. G. Knauff: Diagnosis of the optical properties and structure of lanthanum hexaboride thin films.
J. appl. Phys. **44**, 2252-2256, 1973 (No. 5). *A*

R. C. Peters: Growth characteristics and morphology of GaP liquid phase epitaxy.
Gallium Arsenide and related compounds, Proc. 4th Int. Symp., Boulder 1972, pp. 55-62; 1973. *E*

E. Roeder & A. Troost (Rhein.-Westfäl. Techn. Hochschule Aachen): Einfluß des Stoffverhaltens auf Kraftbedarf und Austrittsgeschwindigkeit beim Strangpressen.
Z. Metallk. **64**, 230-235, 1973 (No. 4). *A*

T. E. Rozzi: Hilbert space approach for the analysis of multimodal transmission lines and discontinuities.
Nato Advanced Study Institute on Network and Signal Theory, Conf. Bournemouth 1972, pp. 289-302; 1973. *E*

T. E. Rozzi: The variational treatment of thick interacting inductive irises.
IEEE Trans. **MTT-21**, 82-88, 1973 (No. 2). *E*

H. Schemmann: Kleine Einphasen-Synchronmotoren mit dauermagnetischem Läufer.
Elektrotechniek **50**, 741-752, 1972 (No. 19). *A*

O. Schob, J. R. de Bie & W. G. A. Klomp (Philips Lighting Division, Eindhoven): The use of television equipment in research on incandescent lamps.
Lighting Res. Technol. **5**, 29-35, 1973 (No. 1).

J. L. Sommerdijk, A. Bril, J. A. de Poorter & R. E. Breemer: Fluorescence decay of Yb^{3+},Er^{3+}-doped compounds, Part I. IR excitation.
Philips Res. Repts. **28**, 475-484, 1973 (No. 5). *E*

R. Spitalnik, M. P. Shaw (Wayne State University, Detroit), **A. Rabier & J. Magarshack:** On the mechanism for microwave amplification in 'supercritically' doped *n*-GaAs.
Appl. Phys. Letters **22**, 162-164, 1973 (No. 4). *L*

F. L. H. M. Stumpers: On the state of the art in information and signal theory.
Nato Advanced Study Institute on Network and Signal Theory, Conf. Bournemouth 1972, pp. 477-493; 1973. *E*

F. L. H. M. Stumpers: Adaptive receiving systems.
Nato Advanced Study Institute on Network and Signal Theory, Conf. Bournemouth 1972, pp. 494-503; 1973. *E*

A. Thayse & J. P. Deschamps: On a theory of discrete functions, Part II. The ring and field structures of discrete functions.
Philips Res. Repts. **28**, 424-465, 1973 (No. 5). *B*

J. B. Theeten, J. L. Domange (E.N.S.C.P., Paris) **& J. P. Hurault:** LEED thermal studies down to very low temperatures.
Surface Sci. **35**, 145-159, 1973. *L*

C. H. F. Velzel: Contours of equal in-plane displacement in holographic interferometry.
Optics Comm. **7**, 302-304, 1973 (No. 4). *E*

J. Vredenbregt & G. Rau (Institute for Perception Research, Eindhoven): Surface electromyography in relation to force, muscle length and endurance.
New developments in electromyography and clinical neurophysiology, editor J. E. Desmedt, Vol. 1, pp. 607-622, 1973.

P. J. de Waard: Measurement of admittance of Gunn diodes in passive and active regions of bias voltage.
Electronics Letters **9**, 59-60, 1973 (No. 3). *E*

K. Walther: ^{55}Mn-nuclear acoustic resonance in $CsMnF_3$.
Phys. Stat. sol. (b) **58**, K 1-3, 1973 (No. 1). *H*

C. Weber: The use of computers in electron- and ion-gun design.
Computer Phys. Comm. **5**, 44-47, 1973 (No. 1). *E*

S. Wittekoek, R. P. van Stapele & A. W. J. Wijma: Optical-absorption spectrum of tetrahedral Fe^{2+} in $CdIn_2S_4$: influence of a weak Jahn-Teller coupling.
Phys. Rev. B **7**, 1667-1677, 1973 (No. 4). *E*

Contents of Philips Telecommunication Review **31**, No. 4, 1973:

H. Kok: Automation in the public telegraph service (pp. 161-173).
H. van Kampen: Telegraph and data switch DS 714 - Mark III (pp. 174-179).
R. J. M. Verbeek: Low-power current mode logic (pp. 180-187).
G. D. M. Vermeire: Application of low-power current mode logic (pp. 188-198).
A. Potuit: Swedish P & T orders Philips 60 MHz line equipment (pp. 200-201).
W. de Vries: HF communications order from Switzerland (pp. 201-202).

Contents of Mullard Technical Communications **13**, No. 121, 1974:

K. W. Stanley & M. H. Dryden: Cleaning processes for Mullard resistors and capacitors on printed-wiring boards (pp. 2-14).
J. Kaashoek: Deflection system design for 110° shadowmask tubes (pp. 15-30).
Phase II horizontal deflection circuit and its design principles (pp. 31-40).

Contents of Valvo Berichte **17**, No. 3, 1973:

J. Siebeneck: Gesichtspunkte zum Entwurf integrierter Mikrowellenschaltungen und Hohlraumoszillatoren für Mikro-Streifenleitungssysteme (pp. 109-120).
W. Schmidt: UHF-Trioden mit hoher Betriebsstabilität in 50 W- bis 200 W-Fernsehumsetzern (pp. 121-128).
W. Schmidt: Entwicklung und Stand der Klystrontechnik für UHF-Fernsehsender (pp. 129-138).

Volume 34, 1974, No. 2/3 pages 33-84 Published 17th June 1974

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 4

Integrated circuits with leads on flexible tape

A. van der Drift, W. G. Gelling and A. Rademakers

Integrated circuits (IC chips) and their various derivatives are expected to form the basic building blocks of electronics for many years to come. No effective competition has as yet appeared, and the quantity production of ICs is particularly suited to automation — computer aids are now widely used in their design and manufacture. The article below presents a solution to the problems of forming the connections between the microscopic crystal chip and the full-sized world outside, in a way that is compatible with automation techniques and the future IC developments for many years ahead. This is a flexible plastic tape, carrying a radial pattern of copper-plated leads for each IC. The technique has been developed by a project group whose members are from the Audio and Video (formerly RGT) and Elcoma Product Divisions and Philips Research Laboratories. Valuable contributions to the work apart from those of the authors were made by F. V. W. ten Bloemendal, H. Budde, H. C. N. van de Sanden and G. J. H. Schermer. The project also benefited throughout from the stimulus and encouragement of Prof. W. van der Hoek. Scientific Adviser, Philips Video Product Division and professor extraordinary at Eindhoven University of Technology.

The integrated circuit and the outside world

In present-day electronic equipment the integrated circuit has assumed a key position. Some equipment is built almost exclusively from combinations of ICs — computer stores, for example — others consist of a combination of ICs and other components. Radio and television receivers are examples of such 'hybrid' circuits, containing many capacitors, inductors, resistors, etc. besides ICs.

In almost all equipment the ICs and other components are mounted on printed-circuit boards carrying the conductors that connect the components to each other and to the outside world. Printed-circuit boards are often made of resin-bonded paper, although recently ceramic plates with thin-film or thick-film wiring have come into use.

It is clear that the integrated circuit has an assured future. It is likely that they will become considerably more complex, so that the problem of connecting them to the outside world will become more and more difficult. The encapsulation now commonly used is relatively large and is not therefore ideally suited to present trends. The encapsulation, necessary to protect the crystal mechanically and from the atmosphere, is also expensive — it often costs more than the integrated circuit itself. This alone was sufficient reason for seeking an alternative using less material and better adapted to automated fabrication, thus leading to lower manufacturing costs.

Fig. 1 shows our solution to the problem. The diagrams give an impression of the technique and of its applications. A flexible plastic tape or film that can be rolled up on a reel is provided with a pattern of plated conducting leads. These leads are connected by soldering to the contact areas of the IC. A special lacquer is applied between the tape and the IC to give protection to this side of the IC. The conducting leads fan out to end in wider areas that can be soldered. Such an IC-on-tape is then moun ted in the circuit by soldering these wider conductor ends to the printed-circuit board. The new technique also offers the possibility of earth-

Ir A. van der Drift is with the Philips Video and Audio Product Divisions, Eindhoven, Dr W. G. Gelling is with Philips Research Laboratories, Eindhoven, and Dr A. Rademakers is with the Philips Electronic Components and Materials Division (Elcoma), Eindhoven.

ing the other side (the back) of the IC crystal chip and of adapting the heat dissipation to suit the requirements. This possibility is significant because of present trends to ICs of higher frequencies and higher powers.

For comparison it is perhaps useful to first look at the conventional way of mounting ICs and connecting them. The main difficulty in the system IC/outside world is the incongruity of dimensions. The contact areas of an IC are separated by a distance of 0.1 to 0.2 mm; the contact holes of a printed-circuit board are at least several millimetres apart. This is a mismatch in dimensions of a factor of about ten. The conventional solution is shown in *fig. 2a* for the case of a Dual-In-Line (DIL) package. In the new technique the mismatch in dimensions is overcome by the radial fan-shaped pattern of conductors on the plastic tape. It is clear that the new technique leads to a reduction in production costs.

Material

1. IC and tape from normal production
2. Value added to IC chip (until final check) is minimal whereas flexibility for final applications is maintained

Technique

1. Can be automated
2. Quantity production

Final product can be mounted

— on printed-circuit boards
— on substrates without holes
— in modified DIL packages

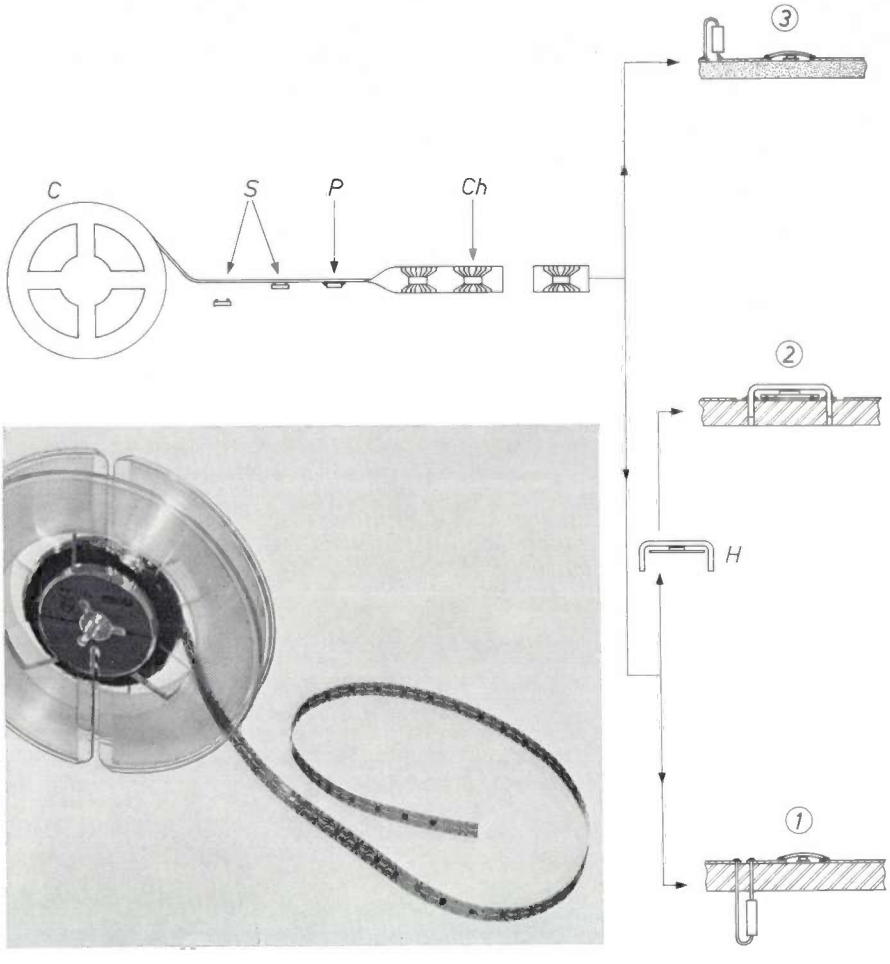


Fig. 1. Summary of the IC-on-tape technique. The photograph shows the new solution to the problem of making connections between an integrated circuit (IC) and the surrounding electronics. The diagrams illustrate the technique and three applications; certain special features are mentioned in the figure. The tape (6 mm wide, 25 μm thick), carrying its patterns of copper conducting leads, is wound on the reel *C*. The integrated circuits *S*, provided with lead-tin solder bumps, are soldered to the leads on the tape. *P* encapsulation (mechanical and atmospheric protection of the active face of the chip). *Ch* testing the individual ICs on the tape. *1*, *2*, *3* examples of applications of the new technique. *1*. An IC-on-tape soldered on a printed-circuit board. The back of the crystal chip is attached to one of the conductors of the printed board with a conducting cement, ensuring electrical contact and some heat dissipation. *2*. Mounting on a printed-circuit board for high heat dissipation. In this version of the IC-on-tape the heat sink *H* (on which the IC chip is cemented) is fixed to the printed-circuit board (see also fig. 12*b* and *c*). *3*. Mounting on a ceramic substrate without holes.

In a DIL package the IC can have between 4 and 48 contact areas, which are connected by very fine gold or aluminium wires to the inner ends of much larger conductors that fan out on a plastic or ceramic base. The complete unit is contained in a block of plastic or ceramic material. Stiff connecting pins extend

out of the block in two rows. These are bent downwards and inserted in the conducting holes of the printed-circuit board to which they are connected by soldering. It is easy to see why the DIL package has become so popular: it ensures protection from mechanical damage or atmospheric corrosion, it allows for some dissipation of heat and it is easy to handle. Yet in recent years there has been a certain amount of re-appraisal: such packages use rather a lot of material and their assembly requires expensive equipment. The whole system is not particularly suitable for automated production. The DIL package has therefore gradually fallen from favour as a good compromise, chiefly because of its relatively large volume and the problems encountered when good heat dissipation is required.

The technique described here does not carry the reduction of size as far as has been done elsewhere, where the whole package is reduced to just the IC chip with a protective layer of glass with solder bumps or beam leads as contacts. Such products achieve the ultimate in size reduction but at the expense of easy handling and easy access for testing, earthing and heat dissipation. The developments often relate to special applications, and are less suited to more conventional uses.

The use of solder bumps is itself a good feature, because it leads to a reduction in production costs: the time-consuming process of making the separate connections is replaced by a short soldering cycle in which all connections can be made at once. This is one of the reasons why in our technique we use solder bumps for the connection of the ICs on the tape.

The plastic film carrying the metal leads is sufficiently flexible to take up small displacements yet stiff enough in its own plane to ensure good positioning of the contacts. Types of plastic are now available that can withstand temperatures up to about 400 °C, so that soldering processes can be performed without damage.

For applying the contact leads to the plastic film we have at our disposal the PD photometallizing process [1] — a technique well adapted to quantity production. The PD process has a very high resolution

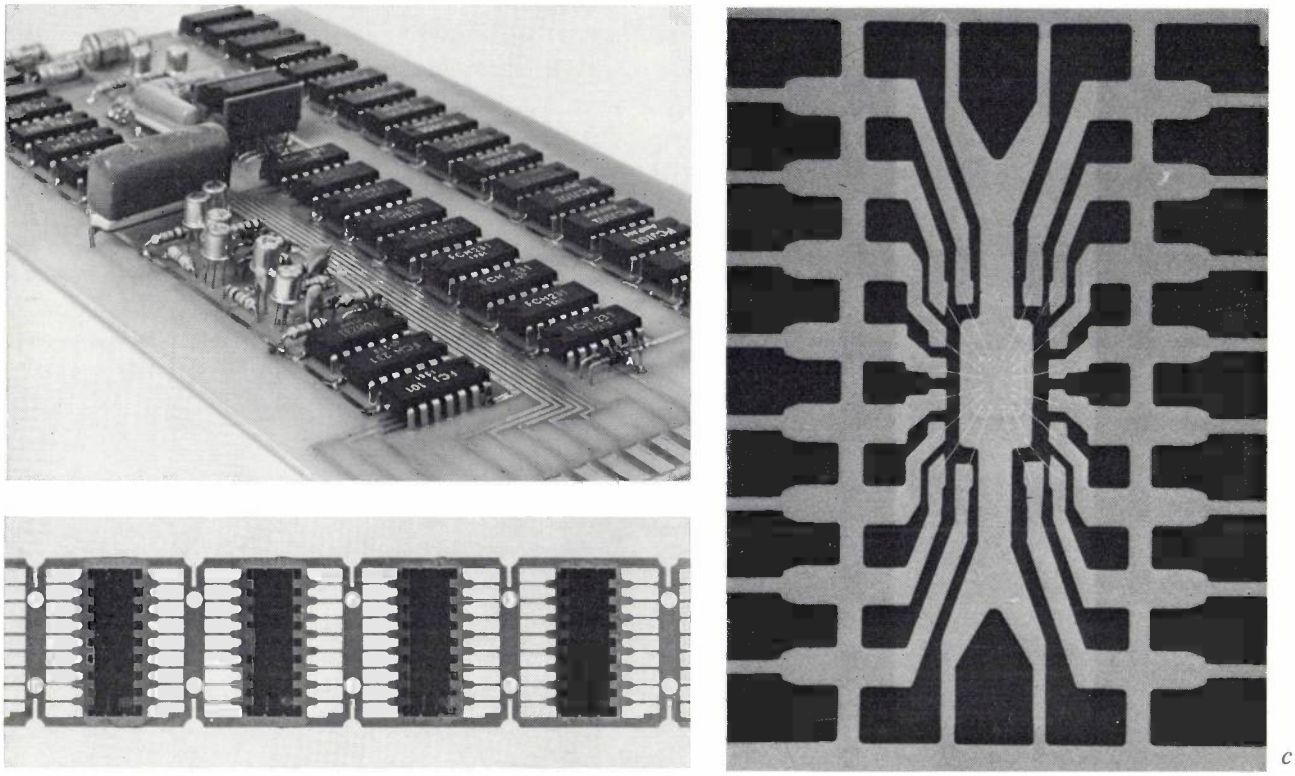


Fig. 2. *a*) The conventional solution for connecting ICs in a circuit. The photograph shows twenty DIL (Dual-In-Line) packages each containing an IC, mounted on a board among other electronic components.
b) A stage in the manufacture of DIL packages showing the (still joined) contact pins. These pins have to fit accurately in the holes of the printed-circuit board (fig. 2*a*); they are then connected by dip soldering.
c) Enlarged X-ray photograph showing an IC mounted in a DIL package. The sixteen microscopic gold wires that connect the IC to the contact pins are just visible. The whole DIL package has a much greater volume than the IC itself.

[1] See L. K. H. van Beek, The PD photographic process, Philips tech. Rev. 33, 1-13, 1973.

and will therefore continue to be effective with the current trend towards still smaller ICs and smaller contact areas.

Metal leads on plastic film have the additional advantage that they can be used as flexible connecting cables. An IC can for example be bonded to a metal base to give a good dissipation of heat while the wide ends of the connection leads can be soldered to another panel some distance away. In this way differences of height can be bridged. Such possibilities imply that the IC-on-tape can be a useful building block in higher-order integration in electronics (multi-level integration), a new and rapidly developing field.

To go one step further, the contact-connection technique described in this article can certainly be extended to include connections to components other than IC chips. Developments may be expected to include contact leads of other metals, finer patterns and plastic films with metal films on both sides (for microstrip lines).

With regard to test measurements on the ICs, the new mounting technique has much to offer. An IC can usually only be tested for correct operation as a finished product. Rejection of an IC therefore involves a smaller loss if its connection and encapsulation is cheaper. A second advantage is that a batch of ICs mounted on a continuous insulating tape are more suitable for automated procedures.

In the manufacture of ICs it is usual to carry out a provisional test at an early stage before the crystal slice is broken up into the separate ICs in order to avoid unnecessary further work on faulty units. The new technique, however, adds so little to the production costs that it becomes profitable to omit this first check (which also adds something to the costs) and to carry out only the final test.

Experience with the new technique of ICs-on-tape has shown that they are very suitable for direct mounting on printed-circuit boards or on ceramic substrates. It is also possible to encapsulate ICs-on-tape in a slightly modified DIL package. Economically, however, such a transitional compromise may become an anachronism and we shall not discuss this possibility further. The new technique will allow electronic equipment to be developed more rapidly and at lower cost partly because conflicting requirements of package design and the accommodation of components will arise less frequently.

In the following sections we shall first discuss the metallizing process for producing the contact leads on the plastic film, and the requirements for the film. We shall then discuss the connection of the IC to the contact leads — the soldering process — and the encapsulation of the IC. The formation of the solder bumps on the IC will be described. Finally we shall discuss the application of ICs-on-tape, in particular the various ways in which they can be mounted in a circuit. Some comments on the complete automation of the new technique will also be given.

Metallizing the plastic tape

The photometallizing machine developed at Philips can operate almost continuously to produce images of the connecting leads in silver on a plastic film. Film with a maximum width of 300 mm and of arbitrary length is fed through the machine, so that it goes continuously through the various stages of the PD process [1]. The film moves at a speed of 180 metres per hour. If we assume the contact-lead pattern of each IC to have the dimensions 15 mm × 6 mm, the machine can produce about half a million pieces per hour, which is more than adequate for present requirements.

The machine is designed for the continuous exposure of the film, which does not have to be stopped, exposed and then transported. A photographic negative (*fig. 3*) carrying an image of the required pattern of contact leads, with a length of about 1200 mm and joined end-to-end to form an endless roll with exact registration of the pattern, is passed round a transparent cylinder and a tension roller. The cylinder rotates continuously, transporting the film, which is held in good optical contact with the negative roll by a rubber idling band. A number of lamps are mounted inside the transparent

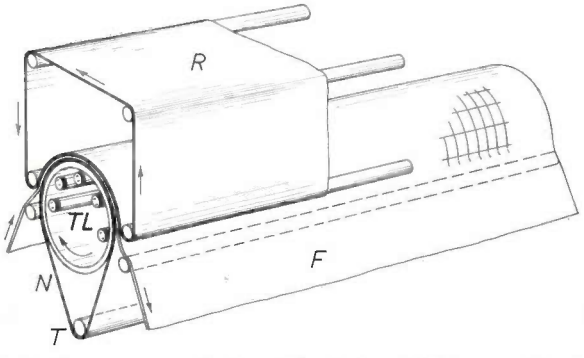


Fig. 3. The exposure cylinder of the photometallizing machine [1]. The negative *N* runs around this transparent acrylic cylinder in the form of an endless roll. The tension roller *T* keeps the negative taut over the cylinder. The film *F* to be exposed is pressed against the negative by the rubber band *R* that idles over four rollers. Inside the cylinder there are a number of fluorescent lamps *TL*. The light from these lamps falls only on the part of the film in contact with the negative. The cylinder is driven continuously to provide the transport of the film. In this way the film is continuously exposed.

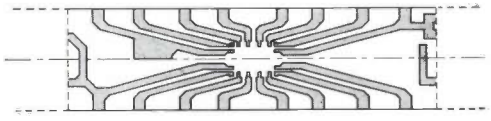


Fig. 4. Layout of the connecting leads on the tape. In this example eight leads terminate on each long side of the rectangle. These wide ends are used to connect the IC to the outside world. The sixteen narrow ends near the centre of the rectangle register with the soldering bumps of an IC. The tape is about 6 mm wide.

cylinder in such a way that only the film in contact with the negative is exposed. The negative roll is easily changed, which is important for the rapid and simple change of production programmes.

Unlike the conventional methods of photography based on silver halides, the PD process gives an image consisting of a *continuous* layer of silver; although this layer is very thin, the image has electrical conductivity. This means that it can be intensified to a much thicker metal layer simply by electroplating. A thicker layer is necessary for our present purpose so that the equipment for the IC-on-tape technique also involves an electroplating machine in addition to the PD machine.

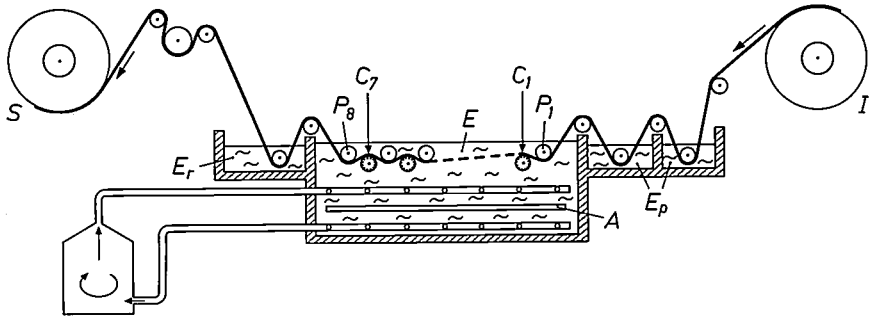


Fig. 5. The electroplating machine shown schematically and much simplified. In this machine the silver images of the conducting leads are intensified with copper, nickel and gold. *E* electrolyte bath. The connecting leads on the film form the cathode. *A* anode. $C_1 \ldots C_7$ rotating contact rollers. $P_1 \ldots P_8$ tension rollers. *I* feed roller loaded with film from the PD photometallizing machine (see also fig. 3). *S* take-up roller. E_p pre-treatment baths for the silver image. E_r rinsing bath; this is followed by drying in hot air. For high-quality electroplating the electrolyte must be well mixed to ensure homogeneity; this is achieved by forced circulation of the electrolyte. This machine was developed by the Electroplating Laboratory of Philips Plastics and Metals Works, Eindhoven.

The plastic film and the connecting leads

The requirements for the plastic film and its pattern of connecting leads are specified primarily by the fact that the IC has to be joined to the connecting leads by *soldering*. The choice of this method of making the connections, which we shall discuss below, is largely determined by the availability of ICs with solder bumps. The simple question of costs naturally implies that the new technique is based on ICs from standard production lines and commercially available types of sheet plastic. The material used for the film is 'Kapton H' [2]. This material can be heated to over 400 °C without chemical decomposition or physical distortion; it can therefore be subjected to the temperatures required for soft-soldering. The film has good dielectric properties and is moisture resistant.

An example of a pattern of connecting leads is shown in *fig. 4*. In this example there are sixteen leads in two rows of eight, ending in relatively wide areas on the two long sides of the rectangle. These ends must register with the contacts of the surrounding circuits. The leads are extremely narrow (0.1 mm) and are very closely spaced near the centre of the pattern. The IC has to be located very accurately in assembly so that the solder bumps lie exactly on these central parts of the leads. For automatic location in assembly the connection-lead pattern must clearly be in exact registration. The endless-roll negative is therefore acceptable only if the image is perfect and repeated in exact registration at the join.

As mentioned above the PD process is followed by electroplating to intensify the image. The complete film can only be electroplated sufficiently uniformly if the individual leads are all electrically interconnected. The image on the negative must therefore also contain the corresponding connections. In the subsequent processing of the film these interconnections are broken or removed. This can be done very simply during the operation in which the whole metallized film is slit into tapes (fig. 1).

The thickness of the plated metal layer is a compromise between conflicting requirements. The resistance of a conducting lead between two contacts should preferably be less than 0.1 Ω. The conducting leads we use are on the average about ten times longer than their width. For copper this implies a thickness of several microns. Also, to give sufficient mechanical strength the copper must not be too thin. On the other hand, flexibility requires that the plated layers should not be too thick. A thickness of about 7 μm gives a good compromise.

Electroplating

The plastic film, now with its rows of silver patterns, comes out of the PD machine on to a roll. The film is then fed continuously to the electroplating machine (*fig. 5*) where the pattern is plated with copper. The

[2] 'Kapton H' (Du Pont de Nemours) is a polypyromellitimide, commonly known as polyimide.

copper leads are next plated with a thin layer of nickel and finally with a thin layer of gold. These two layers protect the copper from corrosion and are indispensable for good 'solderability', as will be shown in the next section. A cross-section of the various layers showing their thicknesses is given in *fig. 6*.

In order to ensure that all the leads and all the patterns are plated uniformly it is essential that the current in the electrolyte should be distributed as uniformly as possible over the film. To achieve this, the anodes are given the form of flat copper plates placed on the bottom of the bath. The cathode is formed by the film with its silver patterns, all the patterns and all the leads being joined together as explained earlier. Electrical contact with these extremely thin silver layers cannot be made with a simple sliding contact since this would damage the image. *Rolling* contacts are therefore used (fig. 5). Adjacent to the contact rollers there are a number of tension rollers that ensure good contact between the film and the contact rollers. The contact rollers span the whole width of the film to ensure a uniform current distribution in this direction. Uniformity of the longitudinal current distribution is ensured by the use of seven closely spaced contact rollers and eight tension rollers over the whole length of the bath.

The contact rollers themselves must *not* of course become plated with metal. To prevent this the contact rollers (*fig. 7*) are made of insulating material carrying twelve inset strips of an alloy of gold and silver. These conducting strips protrude slightly above the surface of the roller. The plastic film runs over the roller in such a way that only the uppermost strips make contact with the silver image. The circuit is connected so that only these uppermost strips carry current. The other strips are therefore not plated. The strips that do carry current at any moment are not plated because of the presence of the film.

The machine shown in fig. 5 is designed for plating with one metal. It is of course possible to carry out all three plating processes for copper, nickel and gold successively in one machine. The current in each plating bath then has to be properly adjusted to match the run-through time in each bath.

Soldering the IC on the tape

Both soldering and ultrasonic welding have been investigated as possible methods for connecting integrated-circuit chips to the connecting leads. Both methods yield reasonably strong connections. Soldering was finally chosen, partly because the inevitable (small) differences in height of the IC contact areas can be taken up better.

The soldering process must satisfy the following conditions:

— the connection must be almost 100% reliable;

— only *soft* soldering (lead-tin or gold-tin) is permissible in view of the maximum permissible temperature of the 'Kapton H' film;

— the soldering should be done without a flux (washing flux away would be difficult and even tiny residues of flux could lead to corrosion);

— the process should be such that all contacts of the IC are connected in *one* operation.

As mentioned in the preceding section the upper layer of the contact leads is of gold, which is easily wetted by molten solder; with copper, direct soldering without flux is more difficult. The intermediate layer of nickel, which is free of oxide owing to the presence of the gold film, is bonded strongly to the solder; the gold dissolves completely in the solder. By making the gold layer sufficiently thin, the formation of intermetallic compounds of gold and tin is prevented; these compounds are brittle and their presence would reduce the strength of the soldered joint. The nickel also has the function of preventing gold loss by diffusion into the copper during soldering; such a loss could cause

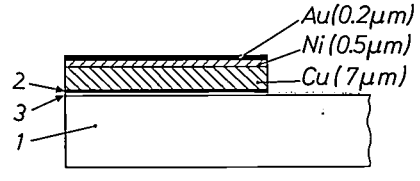


Fig. 6. Cross-section of the film after the electroplating is complete. *1* plastic film ('Kapton H'). *2* silver image made by the PD photometallizing process. *3* cement. The other layers are applied by electroplating. The copper layer provides adequate electrical conductivity and mechanical strength. The nickel and gold layers permit the conducting leads to be soldered with lead-tin solder without flux.

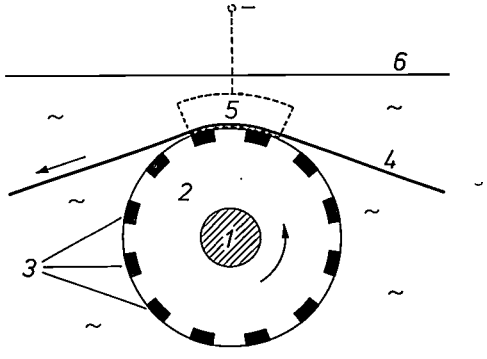


Fig. 7. One of the seven contact rollers in the electroplating machine (cross-section). *1* steel shaft of roller; the arrow shows the direction of rotation. *2* insulating material. *3* gold-silver alloy strips. *4* the film, with the silver image that functions as the cathode; the image is in good contact with the two uppermost strips in the diagram. Only these two strips carry current (via the brush *5*). *6* surface of electrolyte. The silver image is built up uniformly between each two contact rollers. The strips *3* are always shielded by the film while they are carrying current and are therefore not plated.

non-uniform wetting and a poorer connection. The actual soldering operation is only of short duration because of the low heat capacities involved. A protective (forming) gas is used so that little oxidation of the solder bumps takes place during heating. Any oxide layer already present is extremely thin. The pressure on the solder bumps when the IC is placed on the film is therefore sufficient to break this rudimentary oxide layer.

The easy wetting of the gold by the solder could cause the IC to come into too close contact with the tape: it is then very difficult to apply the protective encapsulating layer of lacquer to the surface of the IC chip. For this reason each contact area on the crystal chip is thickened with plated copper, which acts as a spacer and carries the solder bump. *Fig. 8* shows a cross-section of the finished soldered connection.

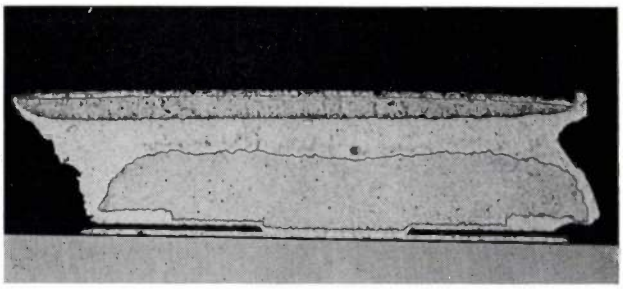


Fig. 8. Cross-section of a solder bump after soldering. The copper spacer, electroplated on the contact area of the IC, can be seen at the bottom (thickness 15 μm, diameter about 100 μm). The solder layer (thickness < 10 μm) that has flowed out over the copper spacer connects a copper lead on the tape (above) to the contact area.

Automatic soldering

A number of experimental machines have been built for the automatic soldering of ICs-on-tape. *Fig. 9* shows an example. This machine operates on a soldering cycle of four stages: (1) feed of tape and IC chips, (2) location of the IC on the tape, (3) light pressure on the IC and tape, (4) heating. The cycle takes about five seconds and is then repeated for the next IC. This machine does not work completely automatically; an operator is necessary to carry out two locating operations which must be checked visually. The feed of the chips is indicated schematically on the left of the figure. A 'pick-up' device places the IC chip with its active side uppermost on an electric heating element (the soldering 'bridge'), where it is held by suction. The solder bumps are on the active upper side of the IC chip. The upper right-hand diagram of fig. 9 shows how the tape is fed from a reel through the machine via two rollers; the side of the tape carrying the connecting leads faces the IC chip. Just above the chip the tape is positioned vertically against a flat surface. Movement of a lever causes the tape to be displaced slightly in a vertical direction. As soon as a good contact is obtained between the solder bumps and the connecting leads the soldering process can start. The required temperature cycle (*fig. 10*) is obtained simply by regulating the current in the heating element, cooling being effected by a stream of gas at room temperature. During the heating, a forming gas is fed to the IC from an array of tubes to prevent oxidation of the solder. To permit the operator to locate the contact leads on the tape accurately with respect to the solder bumps, the soldering bridge carrying the chip is mounted on a carriage that can be moved horizontally in all directions. The tape carrying the soldered ICs is taken up by the left-hand reel in fig. 9.

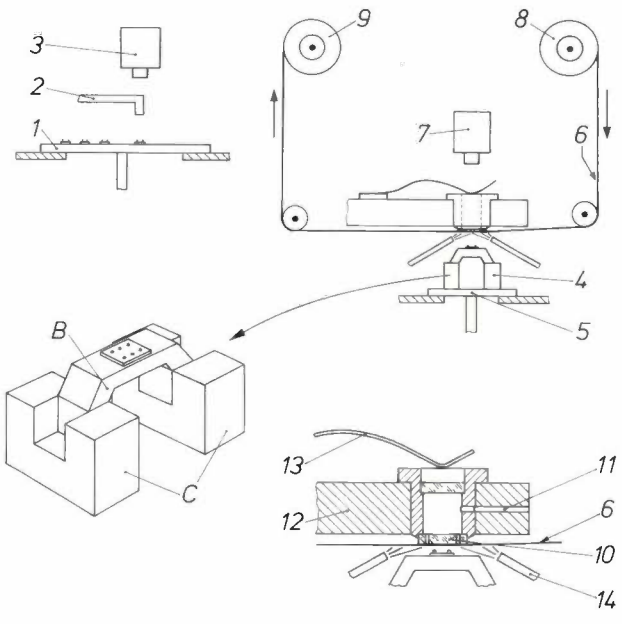


Fig. 9. Diagrams of a soldering machine in which IC chips are soldered to the connecting leads on the plastic tape. *1* moving table for the feed of individual IC chips. *2* pick-up device for transfer of ICs; the operator checks the positioning of the IC in the pick-up by means of the TV camera *3*. *4* soldering bridge; the pick-up places the IC on the heating element of the soldering bridge with its active face upward (see detail drawing on the left; *C* copper contact blocks, *B* electrical heating element). *5* slide for locating the IC (on the soldering bridge) so that its solder bumps lie exactly underneath the corresponding conducting leads on the tape *6*; the operator checks this location on a monitor via the TV camera *7*. *8* feed reel for the tape. *9* take-up reel for the tape with its soldered-on ICs. The tape is held against the glass plate *10* by slight suction from the channel *11*. A small vertical movement of the arm *12* brings the leads on the tape in contact with the solder bumps. Soldering then takes place. *13* spring to ensure the correct contact pressure. *14* array of tubes to supply forming gas and cooling air.

The complete soldering cycle takes very little longer than the actual soldering operation itself because, directly after the soldering of one chip has begun, the next chip is in position to be picked up by the pick-up device. If the cycle were completely automatic, its duration would be about one second shorter; we return to this possibility later. We should also note here that the design of this kind of automatic soldering machine does not depend on the tape being transparent.

Encapsulation

Protection from atmospheric influences — in particular the formation of a water film — is provided by coating the active side of the IC with a special lacquer, which is bonded to the crystal surface by a chemical reaction. During polymerization no gas or vapour is given off: if it were, microscopic leakage paths would certainly occur in the lacquer. Capillary forces ensure that the space between the tape and the face of the IC chip is completely filled with the encapsulating lacquer. An incidental advantage of this is a very considerable reinforcement of the bond between the IC chip and the package. The tape carrying the ICs can be wound up on the take-up reel without any danger of breaking the electrical connections. The lacquer encapsulation is resistant to the high temperatures encountered later in the process.

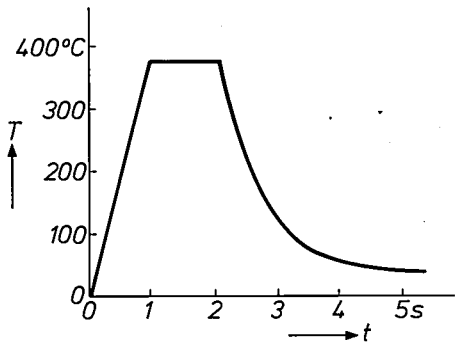


Fig. 10. The temperature T of the heating element (fig. 9) as a function of time t during the soldering of the IC to the connecting leads on the tape. The soldering temperature is reached by passing a current through the heating element for an appropriate length of time. During soldering a forming gas (92% N_2, 8% H_2) is blown against the crystal chip and the tape. Cooling after soldering is accelerated by blowing cool air instead of the forming gas. As soon as the temperature of the chip has dropped to about 100 °C, the tape is transported by one unit so that the following soldering cycle can be started.

Solder bumps

The proper application of the solder bumps with built-in spacers on the IC contact areas is essential to the success of the IC-on-tape technique. If *one* of the 10 000 or so solder bumps on a slice should be missing, then one of the ICs of the perhaps 600 on the slice will be defective. To keep the percentage of defective ICs low, the percentage of missing or faulty solder bumps must not be too large.

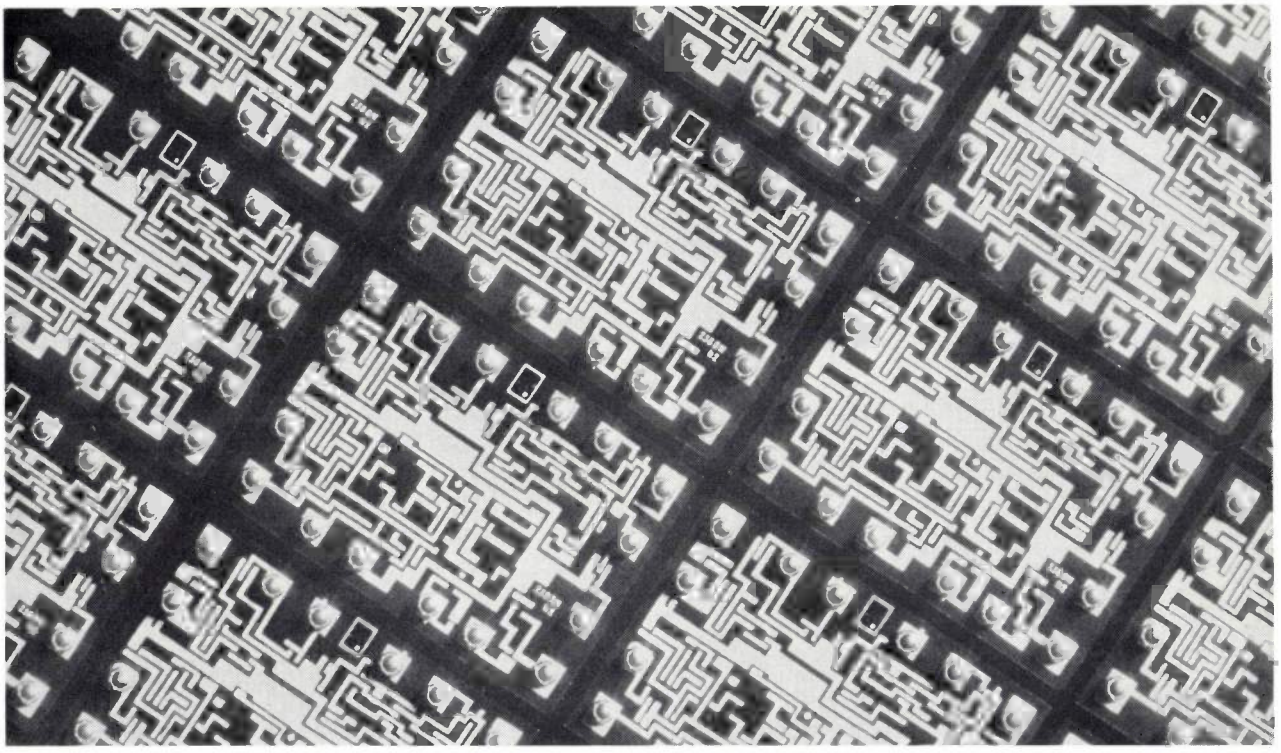
Our process starts with the standard procedures for the preparation of a silicon slice carrying ICs with the usual aluminium metallization. (These ICs can equally well be further processed to DIL packages.) The contact areas are first provided with a layer of nickel by evaporation. The copper spacers are plated electrolytically on to the nickel and the solder bumps (diameter 0.1 mm) are then plated, on to the copper. The nickel coating is necessary because it is rather difficult to plate Cu directly on to Al. The combinations Al-Ni and Cu-Ni, on the other hand, are stable and have sufficiently low contact resistances.

To ensure uniform plating over the whole crystal slice, the local current density must be equal on all the contact areas.This is achieved by means of an auxiliary aluminium layer. The method is as follows. The whole crystal slice with its ICs and aluminium metallization is first coated in a special reactor with a protective layer of glass. The glass is then etched off at the locations of the contact areas. The auxiliary aluminium layer is then evaporated over the whole slice, followed by the nickel layer mentioned earlier. Selective etching is then used to remove the nickel everywhere except at the position of the contact areas. The aluminium layer is not affected by this. The aluminium layer acts as an electrical connection between all the nickel areas so that they all assume the same potential with respect to the electrolyte. The plating of the copper, lead and tin is restricted to the remaining nickel areas by a mask of a thick photographic lacquer with holes at the contact-area locations. When the plating processes have been completed the photographic lacquer and the auxiliary aluminium layer are removed. The thin layer of glass remains. It provides some protection for the circuit and its metallization. A schematic cross-section of an IC on a crystal slice is shown in *fig. 11a*. A photograph of part of a silicon slice with its ICs provided with solder bumps is shown in fig. 11*b*.

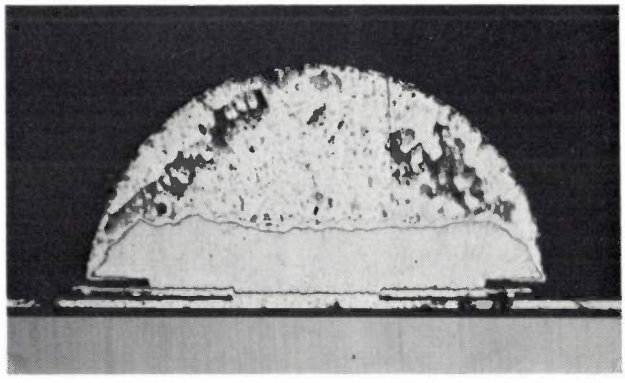
The solder bumps thus far obtained are not hemispherical but mushroom-shaped. The final hemispherical form as shown in fig. 11*c* is obtained by 'remelting' in a bath. The total height of such a bump above the silicon surface of the slice is 60 μm, with a spread of only a few microns.

Mounting in a circuit

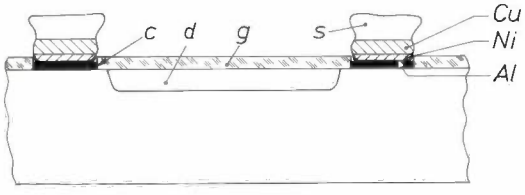
The mounting of an IC-on-tape has to be done in a way that ensures reliable connections between the leads and the surrounding circuit and also in a way that ensures adequate dissipation of the heat developed in the IC chip. In most cases we are concerned with mounting the IC-on-tape on a printed-circuit board. Instead of the robust DIL package we now have a rather flimsy piece of plastic tape with connecting leads on the two edges. The most noteworthy difference is that the IC-on-tape is mounted on the *other* side of the printed-circuit board, i.e. on the side carrying the printed wiring. For each IC there are two sets of matching leads on the printed-circuit board like the teeth of a comb; these provide the connections with the leads on the tape. The teeth of the combs are pre-coated with



b



c



a

Fig. 11. *a*) Schematic cross-section of an IC on a crystal slice. The solder bumps rest on the copper spacers at the locations of the IC contact areas. *d* diffusion region representing the integrated circuit itself. *c* contact area. *g* layer of glass. *Al* auxiliary aluminium layer. *Ni* nickel. *Cu* copper spacer. *s* solder bumps. The aluminium auxiliary film above the glass film has been etched away (see text).
b) Part of a crystal slice. Each IC is provided with sixteen solder bumps.
c) Cross-section of a solder bump. On the contact area of the IC there is an extremely thin layer of nickel. Above this layer there is a layer of electroplated copper and on top of this is the lead-tin solder bump in the form of a hemisphere. The copper layer ensures that a space remains between the IC chip surface and the tape after soldering. This space is necessary to ensure good encapsulation.

solder. There are now a number of possible ways of further assembly, providing for the connection of the ICs-on-tape under all circumstances.

Three cases will be described, all of them referring to mounting on the standard type of printed-circuit board. This limitation, however, is in no way inherent to the methods used.

1. The back of the IC chip is attached to a conductor on the board by means of a conducting cement (e.g. an epoxy adhesive loaded with silver particles). In this way contact is made between the crystal chip and the conductor on the board. This can serve for earthing the IC, which is important at high frequencies. The thermal contact with the conductor ensures, within certain limits, sufficient dissipation of heat. By means of a special tool (a pressure bridge) the connecting leads on the tape are brought into contact with the teeth of the combs on the printed-wiring board. The soldered connection is made by heating the combs just outside the tape (*fig. 12a*).

2. If the heat to be dissipated is considerable, a heat sink is fixed to the back of the chip. The user then only has to fix the heat sink to the board (fig. 12*b*). The heat sink is fixed in place on the board by means of two pins. The connecting leads on the tape are then positioned accurately above the conductors on the printed-circuit board. In this case the conducting leads on the tape do not face those of the printed-circuit board but face away from the board. This has the advantage that the conductors of the printed board can run underneath the IC-on-tape. The IC is soldered to the board by special toothed soldering bits which cause the solder on the conductors of the board to melt just up to the edge of the tape; a particular movement of the soldering bits then makes the solder flow into contact with the leads on the tape (fig. 12*c*).

3. In a third method of mounting a hole is stamped in the printed-circuit board at the place where the IC has to be connected. A heat sink is mounted on the other side of the board and the IC-on-tape — let into the hole — is cemented to this. Compared with the previous method, even more heat can be dissipated; a disadvantage, however, is that the layout of the board is somewhat limited by the presence of the stamped holes.

The above methods of soldering are not restricted to printed-circuit boards; the first method can also be used for connecting ICs-on-tape to ceramic boards with thin-film or thick-film wiring.

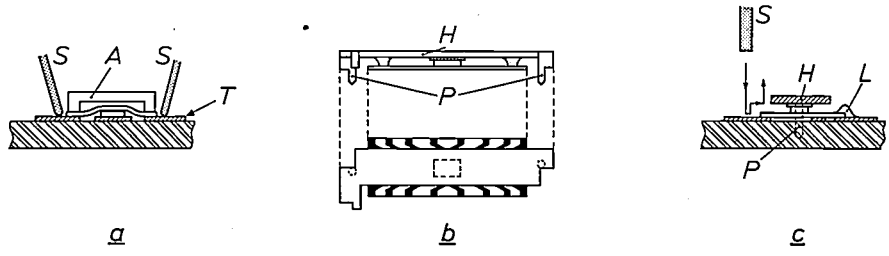


Fig. 12. *a*) Soldering an IC-on-tape to a printed-circuit board. The back of the IC chip is cemented to one of the conducting leads on the printed board by a conducting adhesive. The board carries two connecting combs with solder coatings. The pressure bridge *A* causes the ends of the connecting leads on the tape to make good contact with the teeth *T* of the connecting combs. *S* soldering irons which heat the teeth just outside the tape.
b) An IC-on-tape with the back of the chip cemented to a heat sink *H* by conducting adhesive. This method of mounting gives very good heat dissipation and also allows the crystal to be earthed. The heat sink is fixed to the printed-circuit board on the conductor side, and the leads are then soldered. *P* mounting pins to locate the heat sink accurately with its IC-on-tape so that the connecting leads coincide with the teeth of the connectors on the board. The tape itself is also cemented to the heat sink, in two places. With this method of mounting, the tape with its connecting leads faces away from the board; the conductors of the latter may therefore run under the IC-on-tape (fig. 12*c*).
c) Soldering the connecting leads of the tape to the printed-circuit board in the case of fig. 12*b*. Special toothed soldering bits (*S*, shown only on the left-hand side) are made to follow the path indicated by the arrows so that the solder just outside the tape melts and runs to the connecting leads of the tape. *L* shows the shape of the soldered connection made in this way. *H* heat sink. *P* mounting pins.

Scope for further automation

We saw earlier that complete automation of the soldering process would reduce the time for the whole cycle from five to four seconds. But it is the considerable saving in labour that makes complete automation so attractive. If the visual positioning of the ICs-on-tape were replaced by an automatic method, then the location error of the IC chip on the tape would consist of the sum of four errors. These are: the transport error of the tape, the registration error of the tape, the error in the initial location of the chip and the transport error of the chip. The first two errors are cumulative. When we realize that the solder bumps of diameter 0.1 mm must be located on connecting leads 0.1 mm wide, we conclude that the sum of the errors, at least in one direction, must be less than 20 μm to give acceptable results.

Considering this criterion of 20 μm, we see that the two tape-location errors alone (the transport error and the registration error) are likely to give difficulties unless some feedback control system can be used. A possibility is a system using a repeated registration mark. In this case either an electrical or an optical system appears to be feasible; the connecting leads on the tape have to be located in the machine to an accuracy of typically 10 μm.

The IC chip must be located with equal accuracy; this location refers of course to the solder bumps, and not to the sides of the crystal chip, which, as a result

of the scribing and breaking operation, are no longer located accurately with respect to the solder bumps. There are two possibilities: either the scribing and breaking has to be replaced by sawing, which is much more accurate, followed by location with respect to the edges of the chip; or the use of a reference system with probes that can detect the position of the actual solder bumps.

Both methods make considerable demands on the techniques of high-precision engineering but they are certainly feasible. The automation of the remaining parts of the process will certainly present no greater problems. Our experience with the technique described above therefore leads us to the view that fully automated production of ICs-on-tape is well within reach.

Summary. It is certain that integrated circuits on crystal chips will long remain the most important building blocks of electronics. To simplify and hence cheapen the assembly of complete electronic equipments, a partly automated technique of general applicability has been developed for the electrical connection of ICs to the surrounding circuit. This technique also provides for an adequate protective encapsulation, which takes up very little space. A PD photometallizing machine produces patterns of connecting leads in silver 6 mm in width on plastic film ('Kapton H', 25 μm thick). The film is then fed through an electroplating machine which intensifies the patterns with copper and then adds nickel and gold coatings to allow them to be soldered without flux — this is important to avoid corrosion. The film is then slit into tapes 6 mm in width. A soldering machine then solders the ICs (provided with lead-tin solder bumps) to the inner ends of the connecting leads on the tape. The solder bumps have a built-in copper spacer. This spacer ensures that a capillary space appears between the IC chip and the tape; this space is filled with an encapsulating lacquer to protect the IC and to form a strong bond between the IC chip and the tape. The tape with its ICs is then run through a machine that performs final tests on the ICs. The finished products are delivered in the form of reels of ICs-on-tape.

Several mechanized methods for soldering an IC-on-tape into a circuit, e.g. a printed-circuit board or a substrate without holes are described. In all these cases the IC is mounted on the 'wiring' side. These methods give good heat dissipation and in this respect are decidedly better than ICs mounted in DIL packages. Complete automation of these techniques appears to be a practical possibility.

Philips tech. Rev. 34, 96-102, 1974, No. 4

The compensation wall

P. Hansen and J.-P. Krumme

During investigations on ferrimagnetic materials at Philips Forschungslaboratorium Hamburg, a type of domain wall was found whose properties differ considerably from those of the familiar Bloch wall. One of the most striking properties of the 'compensation wall', as it has been called, is that it does not move when the externally applied magnetic field is increased. Since it has become possible to make magnetic materials in which the compensation walls can be given any desired pattern, there are good prospects for the development of a new type of magneto-optical memory.

Introduction

The state of minimum energy of a magnetic body in the absence of a magnetic field is produced by an inhomogeneous distribution of the directions of magnetization. This state is attributable to the presence of magnetic poles on the surface of the body, which cause a distributed field whose energy can be reduced by magnetic domains of different shapes and different directions of magnetization. In a single crystal the magnetization in the domains points along the 'directions of easy magnetization', as a result of the existence of a magnetic anisotropy field. Provided the demagnetizing field is not too strong, a material with a uniaxial anisotropy field H_u only contains domains in which the magnetization is parallel or opposite to H_u. In a thin platelet with the easy axis perpendicular to the plane of the platelet, elongated domains appear when H_u is greater than the saturation magnetization M_s. *Fig. 1a* shows such domains in yttrium iron gallium garnet, a material we shall take a closer look at in this article. The domains are made visible by a method based on the Faraday effect [1], in which the direction of polarization of an incident beam of light depends on the state of magnetization of the material.

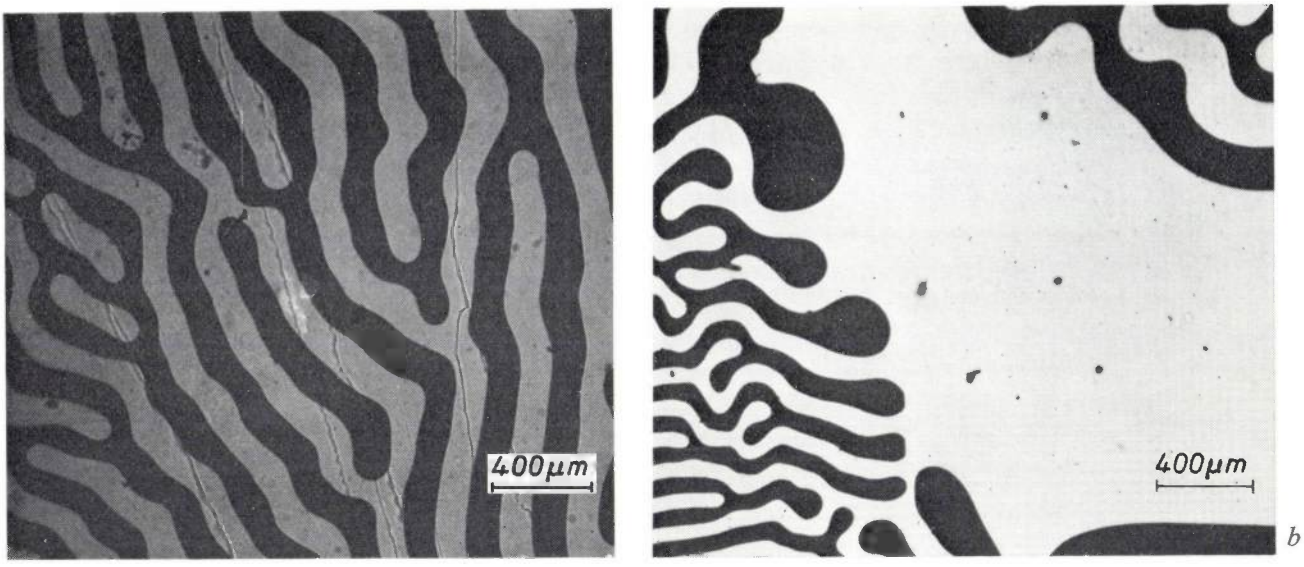
This garnet is ferrimagnetic, which implies that there are at least two sublattices of magnetic ions in the crystalline lattice that have magnetizations of different magnitude and opposite sign (if the sublattice magnetizations are equal in magnitude, so that the resultant magnetization is zero, the material is said to be antiferromagnetic).

The crystals we have studied [2] have the composition $Y_3Fe_{5-t}Ga_tO_{12}$ (YIGaG), in which t varies between 1.2 and 1.3. If the garnet contains no Ga ($t = 0$), one magnetic sublattice is formed by the trivalent Fe ions on the 24 tetrahedral sites and the other sublattice by the trivalent Fe ions on the 16 octahedral sites present in the unit cell (a unit cell contains eight formula units). The resulting magnetic moment per formula unit is therefore the moment of one trivalent Fe ion. At room temperature $M_s = 1.4 \times 10^5$ A/m (or: $4\pi M_s = 1800$ Gs). Substitution of Ga^{3+} ions in the lattice reduces the magnetization because these ions prefer the tetrahedral to the octahedral sites. At $t = 1.3$, which is the concentration in the crystals we have studied, the magnetization is approximately 800 A/m (10 Gs).

Since YIGaG is cubic, a perfect crystal of this material will not have a uniaxial anisotropy field. A uniaxial anisotropy field can nevertheless be obtained, however, under suitably chosen crystal-growth conditions [3] [4]. The domain pattern shown in fig. 1*a* was observed on a homogeneous disc made in this way.

If no special measures are taken during the crystal-growth process, the resultant crystal does not have such a homogeneous Ga concentration as the crystal in fig. 1*a*. In the usual process of growing a crystal from a solution — for example in a lead-oxide melt — the temperature of the solution gradually decreases during the growth of the crystal, so that the gallium content in the crystal varies. In that case t and therefore M_s depend on the position in the crystal, and this in turn affects the domain pattern. The domain pattern in fig. 1*b*, for example, was observed in a platelet in which the magnetization, going from bottom left to top right, first decreases, then passes through zero in the middle of the white part, and finally increases again. This platelet evidently contains a boundary at which the magnetization is zero, which we have called a 'compensation plane' [*]. Without an external magnetic field this boundary cannot be made visible by means of the Faraday effect, because the small gradient of t has hardly any influence on this effect.

Dr P. Hansen and Dr J.-P. Krumme are with Philips Forschungslaboratorium Hamburg GmbH, Hamburg-Stellingen, W. Germany.



a b

Fig. 1. The domain pattern of monocrystalline platelets of the ferrimagnetic garnet $Y_3Fe_{5-t}Ga_tO_{12}$ ($t \approx 1.3$) with a uniaxial anisotropy field perpendicular to the plane of the platelet. The domain pattern is made visible by means of the Faraday effect. *a*) Platelet with homogeneous gallium concentration. *b*) Platelet with a concentration gradient. In the middle of the white part there is a line where the sublattice magnetizations are equal and the total magnetization is therefore zero; this corresponds to the 'compensation plane'.

The compensation wall

The magnetic domains in a homogeneous material, like that in fig. 1*a*, are separated by 'Bloch walls'. Going perpendicularly through the wall from one domain to the next, there is a gradual rotation in the direction of the magnetization of the one domain to that of the other. This is illustrated in *fig. 2*. The distance d_w over which the rotation takes place, called the wall thickness, is determined by the condition of minimum exchange energy and anisotropy energy. In the material to which fig. 1*a* refers, d_w is about 0.3 μm. The inhomogeneous material in fig. 1*b* contains a compensation plane in addition to the Bloch walls. The magnetization, with no applied field, along a line perpendicular to this surface is shown in *fig. 3a*. The resultant magnetization which, as explained in the introduction, is the difference of the two sublattice magnetizations, changes sign in the compensation plane. When a magnetic field H_a is now applied perpendicular to the platelet, one of the Bloch walls present moves to the compensation plane and attaches itself to it. We refer to a Bloch wall that has moved into this situation as a 'compensation wall'. Once it has arrived at the compensation plane the wall remains there, even when the external field is increased in strength. In this it differs from the ordinary Bloch wall which, when the external field is increased, is displaced and finally disappears. This important property of the compensation wall is a consequence of the fact that in a strong external field the magnetization of the sublattices rotates 180° in a compensation plane (see fig. 3*b*) so that the domains are magnetized in the *same* direction on both sides of the compensation wall. Another result of this is that the thickness and energy of a compensation wall depend to a greater extent on H_a [5] [6].

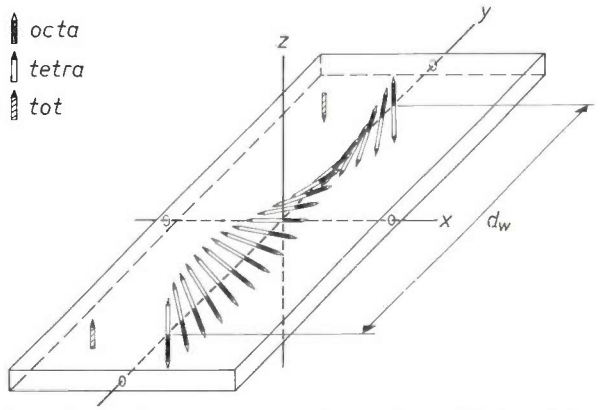


Fig. 2. Illustrating the structure of an ordinary Bloch wall in a ferrimagnetic material. The white and black arrows represent the sublattice magnetizations, the dashed arrows represent the total magnetization. The wall is perpendicular to the *y*-axis.

[1] A description of the optical arrangement has been given in: J.-P. Krumme and J. Haberkamp, Thin Solid Films **13**, 335, 1972, and J.-P. Krumme, P. Hansen and J. Haberkamp, Phys. Stat. sol. (a) **12**, 483, 1972.

[2] The method of preparation has been described by W. Tolksdorf in J. Crystal Growth **3/4**, 463, 1968, and by W. Tolksdorf and F. Welz, J. Crystal Growth **13/14**, 566, 1972.

[3] W. T. Stacy and W. Tolksdorf, AIP Conf. Proc. **5**, 185, 1972.

[4] A. Akselrad and H. Callen, Appl. Phys. Letters **19**, 464, 1971.

[5] P. Hansen and J.-P. Krumme, AIP Conf. Proc. **10**, 423, 1973.

[6] J.-P. Krumme and P. Hansen, Appl. Phys. Letters **22**, 312, 1973.

[*] Not always a true plane in the geometrical sense, but more properly a 'surface'. We use the term 'compensation plane' throughout, however, because it has been established by the authors in their previous publications. (*Ed.*)

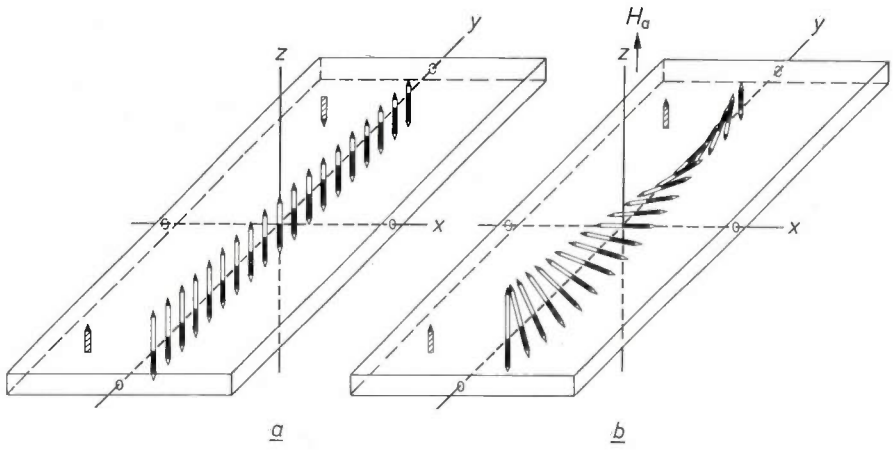


Fig. 3. *a*) Sublattice magnetizations and total magnetization (see fig. 2) in the vicinity of a compensation plane ($y = 0$) along a line perpendicular to this plane in the absence of an external magnetic field. *b*) As (*a*), in the presence of a magnetic field H_a after the formation of a 'compensation wall'.

Another property that distinguishes the compensation wall from an ordinary Bloch wall is that the compensation wall can be displaced by a change of temperature. This is due to the different temperature dependences of the sublattice magnetizations, and therefore a change of temperature causes a shift in the compensation plane and hence in the position of the wall. We have observed a displacement of no less than 80 μm per degree [5].

Fig. 4a shows another garnet platelet like that in fig. 1*b*, but now containing a compensation plane in addition to Bloch walls. There is a weak magnetic field perpendicular to the plane of the platelet, and there is a compensation wall (indicated by the arrow) in the compensation plane. The thickness of both types of wall is considerably less than 1 μm. Fig. 4*b* shows a roughly circular compensation wall, referred to as a 'compensation bubble'. This configuration will be discussed in more detail below.

In the foregoing we have confined the discussion to vertical, planar and cylindrical compensation planes, but in general a compensation plane in a given sample can have any position and shape. In the cases considered, a weak field is sufficient for a Bloch wall to attach itself to the compensation plane. Generally speaking, however, the situation is different [7] [8]. Let us look, for example, at a thin platelet whose normal is at an angle to the normal of a planar compensation plane inside it. The existence of a compensation wall in this situation is illustrated schematically in *fig. 5*. In a weak external field H_a perpendicular to the platelet there is a straight Bloch wall in the middle of the compensation plane (fig. 5*a*). As the field-strength increases, part of the wall attaches itself to the compensation plane and that part of it become a compensation wall (fig. 5*b*). In the limiting case of very strong fields the compensation wall extends along the entire compensation plane (fig. 5*c*). This situation is confirmed by measurements of the local Faraday rotation of a beam of light projected perpendicularly on to the platelet. The relative rotation Θ/Θ_0 is shown in *fig. 6*, where Θ_0 is the Faradary rotation of the uniformly magnetized platelet.

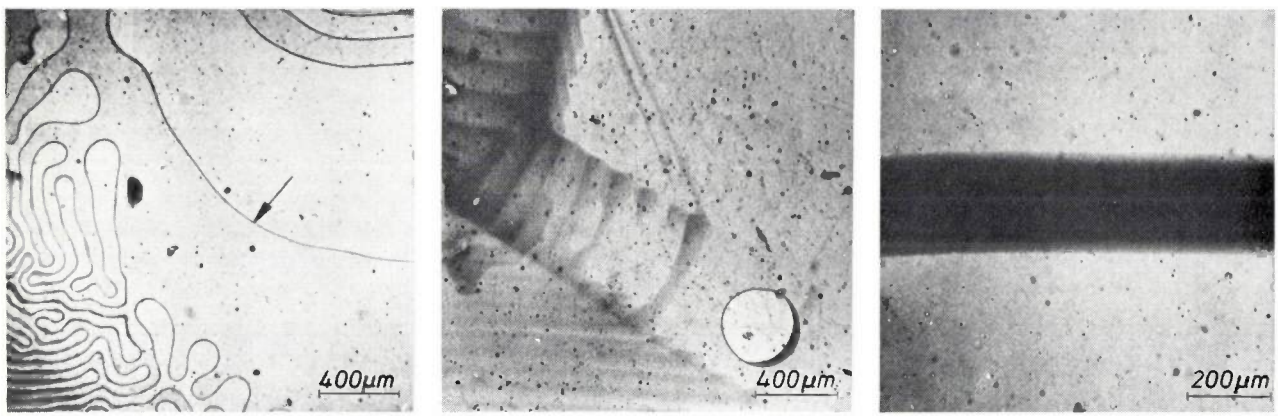


Fig. 4. Pictures of compensation walls in the garnet platelet in fig. 1*b* (thickness 90 μm). *a*) A part which contains a compensation wall (see arrow) in addition to ordinary Bloch walls. *b*) Closed compensation wall in an external field H_a of 1600 A/m (20 Oe). *c*) Inclined compensation wall in a field H_a of 17 000 A/m (213 Oe). At this field-strength the projection of the wall is no less than 250 μm wide.

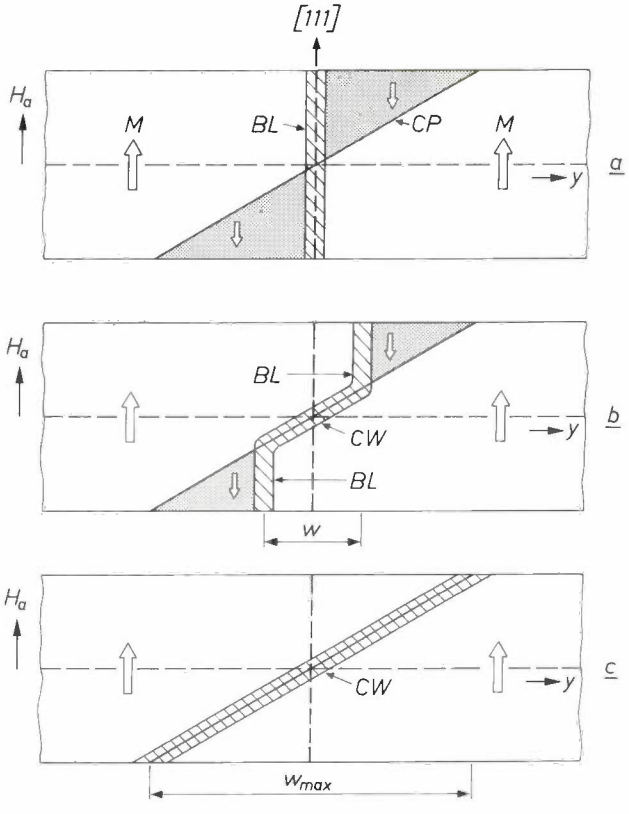


Fig. 5. The formation of a compensation wall along an inclined compensation plane under the influence of an increasing magnetic field H_a perpendicular to the platelet. *CP* compensation plane. *Bl* Bloch wall. *CW* compensation wall. *M* is the total magnetization and w and w_{max} are respectively the width and maximum width of the projection of the compensation wall. *a*) At low values of H_a a Bloch wall is situated in the middle of the compensation plane. *b*) With increasing field-strength part of the Bloch wall attaches itself to the compensation plane and there becomes a compensation wall. *c*) At high field-strengths the compensation wall extends along the entire compensation plane ($w = w_{max}$) and w cannot increase upon a further increase of H_a.

The symbol w_{max} stands for the maximum projection of the inclined compensation plane on the plane of the platelet. The field-dependence of the projection w is shown in *fig. 7*. The slope of the compensation plane can be derived from the value of w_{max} and the thickness of the platelet. In the pictures shown in fig. 4, the process described here appears as an apparent thickening of the wall (see fig. 4*c*).

The compensation bubble

Fig. 4*b* gives an example of a radial gradient of the gallium content in a YIGaG disc, which resulted in a cylindrical compensation wall [9]. The position of a 'compensation bubble' remains unchanged when the field changes, but the cross-section of the bubble can be changed by a change in temperature. A compensation bubble can be formed from a cylindrical Bloch wall that surrounds the area with the radial gallium gradient and whose cross-section decreases with increasing magnetic field until the wall reaches the cylindrical compensation surface. The shape and position of a compensation bubble are determined by the local variation of the compensation temperature T_{comp}, i.e.

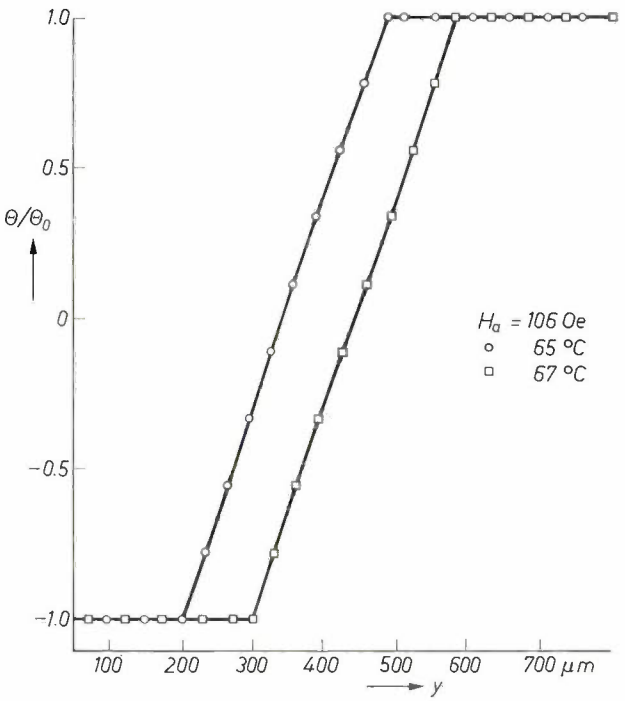


Fig. 6. Faraday rotation Θ along a path y perpendicular to the line intersecting the surface of the garnet platelet and the inclined planar compensation wall, normalized to the maximum rotation Θ_0 of the uniformly magnetized disc. The measurement was made in a field of 8500 A/m (106 Oe) using a focused laser beam (diameter 5 μm) at two different temperatures. As can be seen, Θ/Θ_0 in the region of the wall is linear in y, and on a temperature change of 2 °C the wall moves through a distance of 100 μm.

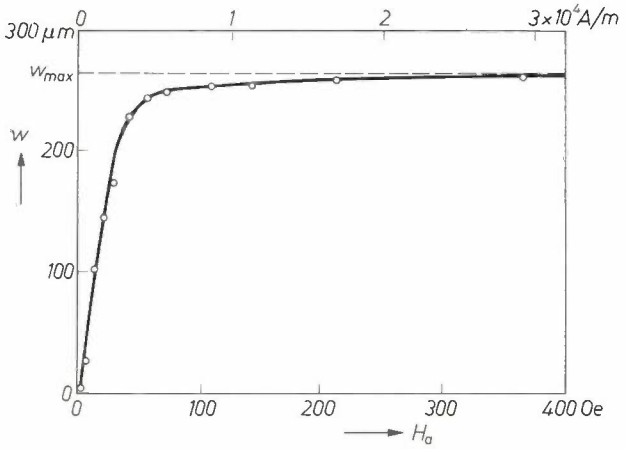


Fig. 7. Field-dependence of the width w of the projection of the inclined compensation wall. At low field-strengths, w depends strongly on the field H_a. At high field-strengths w approaches a maximum value w_{max}.

[7] R. C. Le Craw and R. Wolfe, Proc. Conf. on Magnetism and Magnetic Materials, Boston 1973, paper No. 4B-7.
[8] J.-P. Krumme, Phys. Stat. sol. (a) **23**, 33, 1974.
[9] J.-P. Krumme and P. Hansen, J. appl. Phys. **44**, 3805, 1973.

the temperature at which the local magnetization becomes zero. In *fig. 8* the value of T_{comp} is plotted as a function of a coordinate y that lies in the plane of the platelet and passes through the centre ($y = 0$) of the

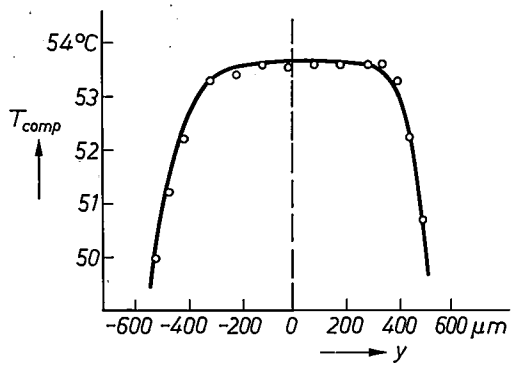


Fig. 8. The temperature T_{comp}, the 'compensation temperature', at which the sublattice magnetizations are identical, measured along a straight line (coordinate y) through the centre ($y = 0$) of a compensation bubble.

existence region of the compensation bubble. The temperature T_{comp} was measured magneto-optically by locally determining the temperature at which the hysteresis loop reverses. The size and location of the compensation bubble can be read from fig. 8 for a given temperature. The compensation bubble cannot exist at temperatures higher than the value of T_{comp} in the centre.

Applications

The practical application of compensation walls implies the ability to give any desired shape to these walls in a ferrimagnetic material. This has become possible with a recently devised method [10] for locally controlling the saturation magnetization of iron-garnet films with a high content of diamagnetic ions, such as Ga^{3+} and Al^{3+}, on tetrahedral sites and partly on octahedral sites as well. The method consists in depositing a silicon film about 200 nm thick on a garnet film grown epitaxially from the liquid phase (liquid-phase epitaxy, LPE), and then subjecting it to a heat treatment at about 600 °C. In the silicon-covered parts of the garnet film the saturation magnetization and the compensation temperature then undergo a change, presumably as a result of a change in the distribution of the diamagnetic ions among the sublattices.

With this method we have made various patterns of compensation walls that can be used for memories in which information is written in thermomagnetically and read out magneto-optically [11]. In a garnet film of the composition $Y_{2.34}Gd_{0.52}Yb_{0.14}Fe_{3.74}Ga_{1.26}O_{12}$ we have produced a rectangular pattern of the magnetization and the compensation temperature, which is explained schematically in *fig. 9*. In the areas treated with silicon (x_1,x_2) the compensation temperature has become higher as a result of the increased magnetization of the tetrahedral sublattice and the decreased magnetization of the octahedral sublattice.

If an ambient temperature T_3 between the compensation temperatures T_{comp1} of the untreated part of the garnet film and T_{comp2} of the treated part is chosen, and a magnetic field H_a is applied perpendicular to the film, compensation walls are then formed at the positions x_1 and x_2 (fig. 9). The direction of the octahedral-sublattice magnetization, which is predominantly responsible for the Faraday rotation, is shown for this temperature in fig. 9*c*. The pattern of the rectangular compensation-wall domains can be seen in *fig. 10* for applied fields H_a of various strengths and for various temperatures. It is observed that the walls only lie along the compensation planes when the field-strengths

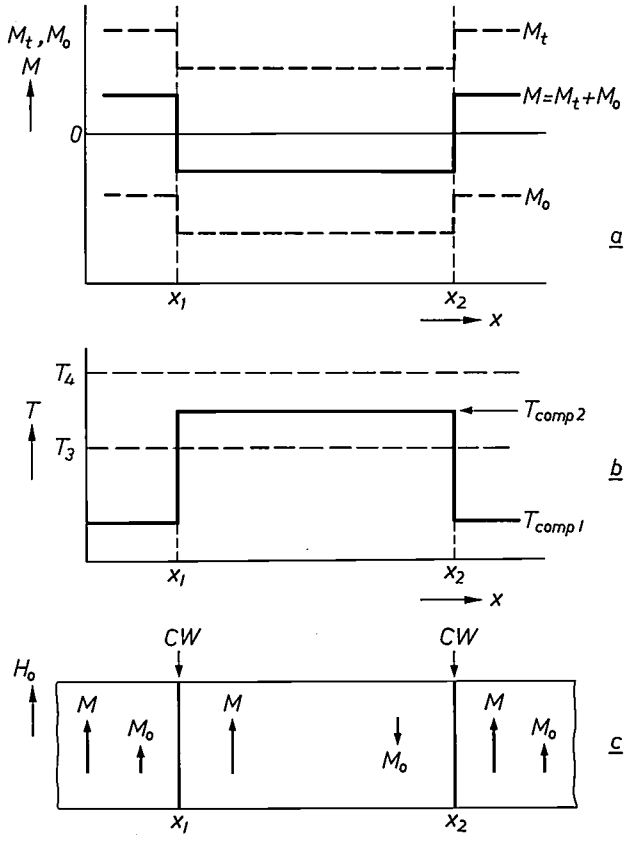
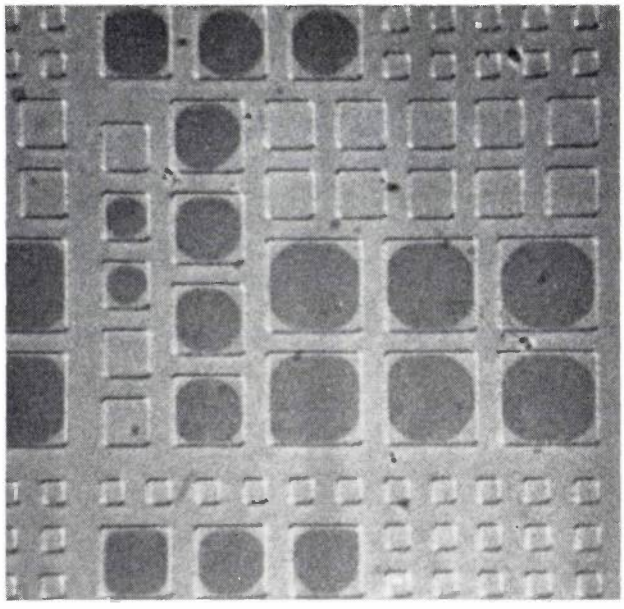
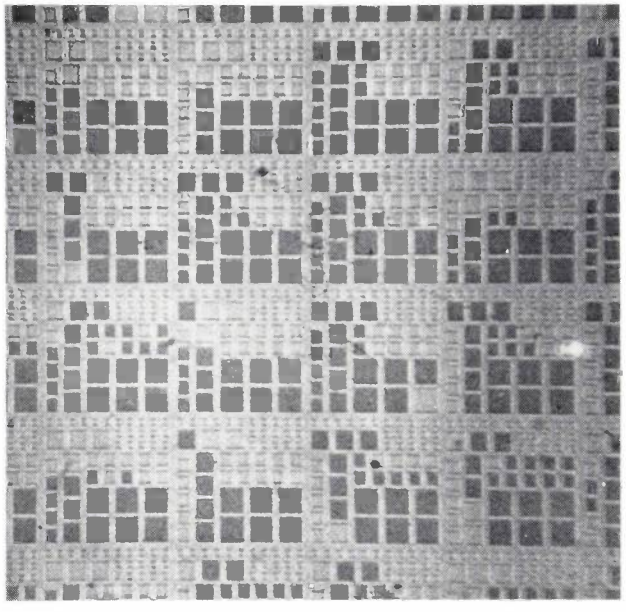


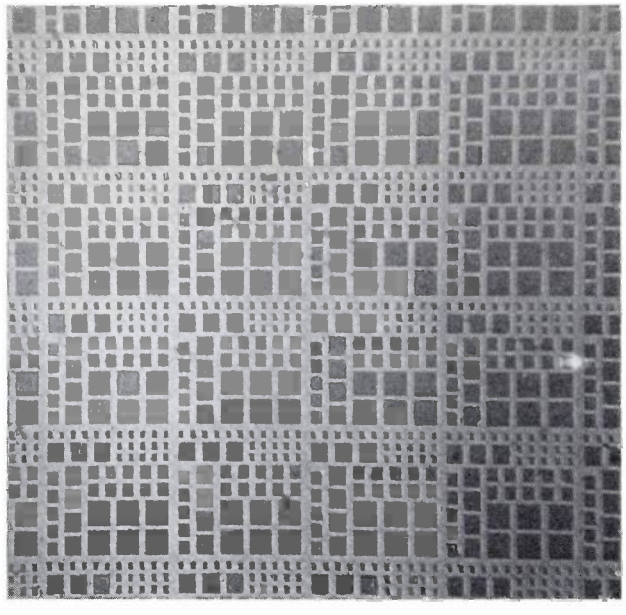
Fig. 9. The saturation magnetization can be locally controlled by covering parts of the surface of a YFeGa garnet platelet with silicon and then subjecting the surface to a heat treatment. In this way a pattern of compensation walls can be produced in the disc. *a*) Local variation of the tetrahedral magnetization M_t, the octahedral magnetization M_o and the total magnetization M. At the limits x_1 and x_2 of the part treated with silicon the magnetization changes sign at a given temperature. *b*) In the part between x_1 and x_2 the compensation temperature T_{comp} is higher than in the untreated part. When the temperature T_3 of the crystal lies between T_{comp1} and T_{comp2}, the effect of an applied magnetic field is to produce compensation walls at x_1 and x_2. *c*) The direction of the magnetizations M and M_o in the region (x_1,x_2) between the compensation walls at CW and in the surrounding area.



a



b



c

Fig. 10. Garnet platelet ($Y_{2.34}Gd_{0.52}Yb_{0.14}Fe_{3.74}Ga_{1.26}O_{12}$) in which a regular pattern of compensation walls has been produced. The three photographs relate to situations in which the external field H_a and the temperature have different values. *a*) $H_a = 2400$ A/m (30 Oe), the temperature is 5 °C lower than T_{comp2} (see fig. 9). As can be seen, only the larger domains show a reversal, and the dark patches do not extend to the corners of the square fields. *b*) $H_a = 4000$ A/m (50 Oe), the temperature is 7 °C lower than T_{comp2}. The fields are now dark right up to the corners of the squares, and some of the smaller domains have also undergone a reversal. *c*) $H_a = 8000$ A/m (100 Oe), the temperature is 19 °C lower than T_{comp2}. All the domains, even the smallest, have reversed.

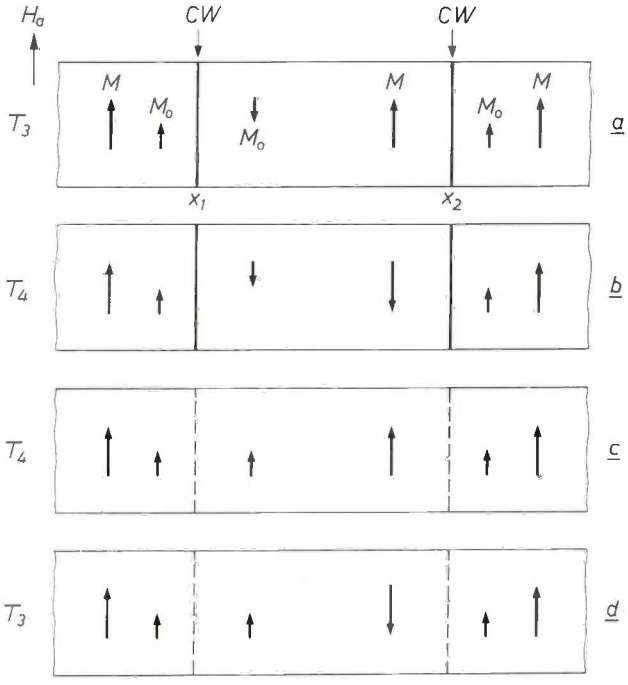


Fig. 11. Illustrating the cyclic thermomagnetic recording process in a memory based on compensation-wall domains (see fig. 9*c*). The symbols used have the same significance as in fig. 9.

are relatively high, and that the formation of small compensation-wall domains by field reversal is only possible at temperatures that are not too close to T_{comp2}.

The switching cycle of a compensation-wall domain is illustrated in *fig. 11*. In the starting situation (*a*) we have a pattern of compensation walls. When the temperature in the region (x_1,x_2) is locally increased to the temperature T_4, which is slightly above T_{comp2}, the sign of the magnetization reverses (*b*). The new situation is not stable in sufficiently strong fields; the magnetization reverses and the walls disappear (*c*). When the film has cooled to the ambient temperature T_3, the situation

[10] R. C. Le Craw, P. A. Byrnes Jr., W. A. Johnson, H. J. Levinstein, J. W. Nielsen, R. R. Spiwak and R. Wolfe, IEEE Trans. **MAG-9**, 422, 1973.

[11] J.-P. Krumme and P. Hansen, Appl. Phys. Letters **23**, 576, 1973.
J.-P. Krumme, W. Tolksdorf, H. Dimigen and H. Hieber, Phys. Stat. sol. (a) **20**, 725, 1973.

d is found. Because of the threshold energy for the formation of domain walls at the positions x_1 and x_2, this new situation is stable. The starting situation (a) can be obtained again by applying an additional field pulse which realigns the magnetization in the region (x_1,x_2) parallel to the field.

Switching processes as described above have been carried out in a constant magnetic field H_a of only 3280 A/m (41 Oe). Pulses of light of 5 microseconds duration from a continuous-wave argon laser of 25 mW (wavelength 5145 Å) were sufficient to heat up the material in a compensation-wall domain of 12×12 μm in a layer 3 μm thick. The beam diameter was approximately 15 μm and the temperature of the layer was kept 16 °C below T_{comp2}. It is believed that in the long run a switching energy of less than 3×10^{-3} erg/μm^2 will be sufficient for this technique.

From the results so far achieved it may be concluded that this new magneto-optical memory is a promising alternative to conventional types of magneto-optical memories.

Summary. In a crystal platelet of Ga-substituted yttrium iron garnet ($Y_3Fe_{5-t}Ga_tO_{12}$ with $t \approx 1.3$) a new type of magnetic 180° wall has been found, which has been called the 'compensation wall'. Platelets of this ferrimagnetic material in which there is a gradient in the Ga concentration may contain planes in which the magnetizations of the sublattices compensate one another. When a magnetic field is applied, a Bloch wall moving towards such a plane does not go beyond it, but forms a compensation wall there. The wall cannot be displaced by varying the magnetic field, but it *can* be moved by varying the temperature. The diameter of a closed (cylindrical) compensation wall, the 'compensation bubble', varies with temperature. If the compensation plane is inclined, the Bloch wall settles towards this plane as the magnetic field increases. It is possible to produce regular patterns of compensation walls in garnet films obtained by liquid-phase epitaxy (LPE). These films can form the basis for a new type of magneto-optical memory.

Grease-lubricated helical-groove bearings of plastic

Two of the many attractive features of the helical-groove bearing [*] [1] [2] with grease as the lubricant are that the rotating shaft is supported by a continuous film of lubricant (full-film lubrication), and also that when certain conditions are fulfilled, no seals are required to keep the lubricant in the bearing. This is because the helical grooves act as a hydrodynamic seal when the bearing is running, whilst the consistency of the grease prevents the lubricant from escaping when the bearing is stationary.

The full-film lubrication leads directly to a number of technical advantages. When the shaft rotates, there is no contact between the bearing surfaces; the bearings are therefore quiet, and have a constant frictional torque, a constant centre-line position when statically loaded, a greater load-carrying capacity than porous oil-impregnated sintered-metal bearings and a long life. Disadvantages of helical-groove bearings are that the bearing has only one permissible direction of rotation, tolerances have to be small and its speed must be greater than a certain critical value. (Below this speed full-film lubrication cannot be achieved.)

Apart from their use in professional equipment, helical bearings are very attractive for electric motors and other components in mass-produced articles in which servicing or oiling is undesirable or impossible, as in domestic and audio equipment. This underlines the present-day tendency to design equipment with few maintenance problems and long life. Good bearings are especially important in audio equipment: the quality of the sound is greatly dependent on the quality of the bearings.

With mass-produced articles there is, apart from the functional criteria, the requirement that the product should be cheap and easy to manufacture in large numbers. For the mass production of helical-groove bearings, a likely possibility would seem to be the use of injection-moulded plastics, provided that the heat developed in the bearing while running does not exclude the use of a relatively poor conductor of heat. Moreover it is essential that the injection-moulding technique should be capable of sufficiently accurate control to give the necessary small tolerances.

The bearings mostly used nowadays in mass-produced consumer articles, porous oil-impregnated sintered metal bearings, have been in use for many years, and much experience of them has been accumulated. Before a production department changes such an essential component as a bearing and replaces it by a completely new device such as a helical-groove bearing, the usefulness and the reliability of such a new bearing must first be very carefully demonstrated by life tests.

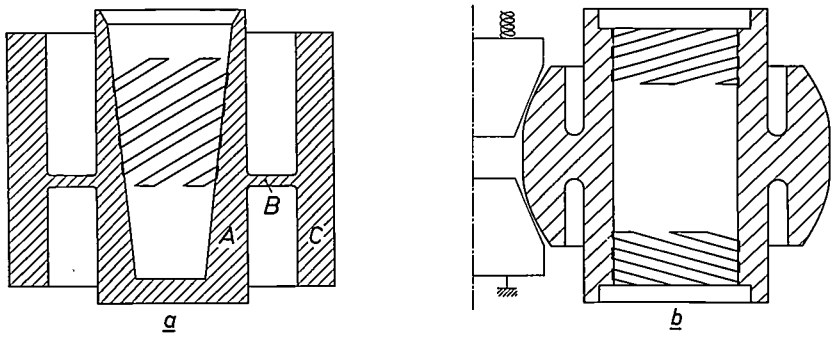


Fig. 1. The two types of plastic helical-groove bearing used for the life-tests. *a*) Blind conical bearings consisting of the bearing bush proper *A* with its internal helical grooves, supported by flexible 'spokes' *B* in an outer mounting ring *C*. In the first group of bearings to be tested, the ends of the spokes *B* were not rounded. *b*) Journal bearing with self-aligning spherical mounting.

We have therefore subjected various types of helical-groove bearings to such tests, first of all in batches of ten and later in batches of twenty. The results, reported below, reveal clearly the potentialities of these bearings. The first life tests (begun in 1966) were with blind conical 'helical-groove' bearings made of plastic and mounted in asynchronous motors of 30 W running at about 2850 rev/min (shaft diameter 5 mm, bearing load about 10 N). The bearings were blind conical helical-groove bearings (*fig. 1a*); this type of bearing was chosen for the first tests because it can be fabricated using a very simple injection mould and because, with

[*] These bearings are sometimes referred to as 'spiral-groove' bearings.

[1] E. A. Muijderman, Philips tech. Rev. 25, 253, 1963/64.
[2] G. Remmers, Philips tech. Rev. 27, 107, 1966.

a blind bearing, the lubricant can only escape on one side.

Mounting of the bearing in the motor frame is shown in *fig. 2*. A flexible support provided by 'spokes' ensures that the bearing is well aligned and takes up any thermal expansion of the motor shaft. The load was applied via a lever with a rubber-coated idler wheel; the idler wheel showed non-uniform wear on its circumference after some time so that the load on the bearing was irregular (chatter).

The bearings investigated were not all exactly alike, but modified versions of the type shown in fig. 1*a* were also investigated, such as bearings with three or six spokes. (It was found that three spokes were too few; fatigue cracks soon became apparent.) Another design was also tried out in which the flexible spokes were replaced by a spherical pivot on a flexible membrane; this arrangement was also satisfactory although it was more sensitive to mounting errors and, because it consisted of *two* components, it was more expensive than the arrangement with spokes. Finally, for comparison, a number of bearings were made in which the helical grooves were omitted: the life-test results for these bearings showed very convincingly the vital function of the grooves.

The material for the grooved plastic bush was chosen after an investigation of the dry-running behaviour of various plastics against the shaft steel. A polyacetate copolymer was finally chosen. The lubricant was a semi-fluid grease with a highly viscous mineral-base oil and a sodium-soap thickener.

A year after the above tests were begun, life tests were started with eighteen motors with six flexible spokes that had *rounded* ends (fig. 1*a*). Eight of these motors were run in a start/stop cycle (half-minute on, half-minute off), and after about 500 000 start/stops they were run continuously like the others. Six of the eighteen motors have meanwhile failed after running times of 30 500, 36 500, 37 500, 42 500, 47 000 and 48 500 hours; the other twelve are still running (5 of these for 45 000 hours with 500 000 start/stops; the other seven for 50 500 hours).

Statistical analysis of these life-test results, together with extrapolations of the results of previous life-tests, gives the survival curve *a* shown in *fig. 3*. To give a comparison of the results of the helical-groove bearings with those of motors using conventional sintered-bronze bearings, ten of these motors were life-tested under the same conditions. These tests gave the curve *b* in fig. 3. Comparison of the two survival curves shows immediately that the helical-groove curve represents a more reliable bearing. For a long time the failure rate of the polyacetate helical-groove bearing is practically zero; then the failure rate rises fairly rapidly [3]. The

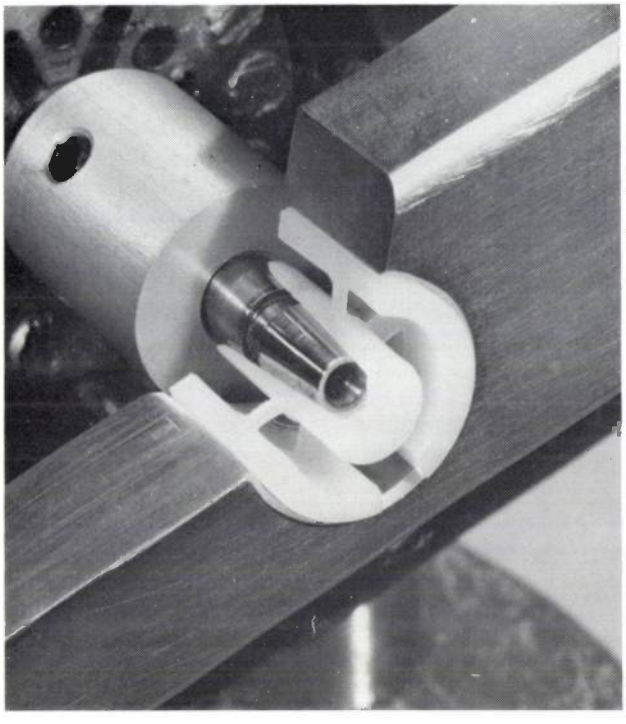


Fig. 2. Photograph of a blind conical helical-groove bearing of plastic mounted in an electric motor. The outer ring and the bearing bush are cut away to show the flexible-spoke mounting of the bearing bush itself. The outer plastic ring (*C* in fig. 1) is locked in position by a screw that expands the slot just visible under the bearing bush.

motors with sintered-bronze bearings have the highest failure rate right at the start of operation. The life achieved by 90% of the motors is therefore about 10 times higher with the helical-groove bearings than with the sintered-bronze bearings: 36 000 against 2600 hours. The good performance of the plastic helical-groove bearings is undoubtedly the result of the hydrodynamic lubricating film maintained so easily by the grooves. The presence of full-film lubrication was confirmed by measurements of the electrical resistance between the shaft and a metal helical-groove bush identical to the plastic version.

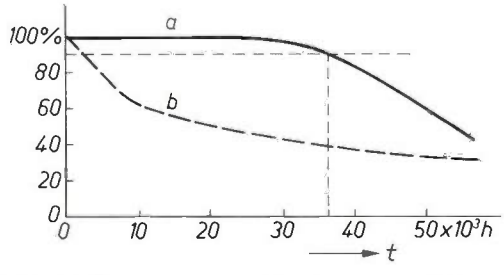


Fig. 3. Survival curves representing the results of life-tests on electric motors fitted with helical-groove bearings of plastic. Curve *a* refers to tests on 18 motors fitted with blind bearings of the conical type shown in fig. 1*a*. Curve *b* refers to tests on ten motors fitted with conventional porous sintered-bronze bearings. The ordinate represents the probability that a motor is still running after the time *t* shown along the abscissa. The horizontal line at 90% intersects the curves at points which represent the life achieved by 90% of the motors [3].

Further life tests were made on helical-groove bearings of the journal type (fig. 1*b*). Helical grooves in a cylindrical bearing of this type require a more complicated injection mould. Because thermal expansion of the motor in the direction of the shaft no longer has any effect on the bearing clearance the spoke mounting is no longer necessary and self-alignment can be achieved simply by making the outside of the bearing spherical, exactly like the mount used for sintered-bronze bearings.

After the tests described had been running for some time, the growing interest in plastic helical-groove bearings lead to the extension and intensification of the test programme for the journal bearings. The criterion used to decide whether a motor was to be considered as failed is the motor speed (rev/min); if this deviates by more than 2% from the initial value, the motor is rejected. Moreover, the life tests were now carried out at two ambient temperatures: 25 °C and 45 °C. This second series of life tests was started with a group of 10 motors (5 at 25 °C and 5 at 45 °C); these motors have meanwhile run for 4000 hours with no failures so far.

After this life test had run about 1500 hours and continued to indicate good prospects of life, the tests were extended with 30 motors, 20 fitted with plastic helical-groove bearings and 10 with sintered-bronze bearings. Measurements of the noise level were made on all these motors. It was found that there was initially little difference in the noise level between the two groups of motors. After 2000 hours of running, however, it was found that the noise level of the plastic helical-groove bearings had dropped slightly and that there was less scatter among the various test motors whereas, for the sintered-bronze bearings, the noise level was slightly higher and the scatter slightly greater.

Although these tests are still under way, it is already clear that helical-groove bearings made of plastic represent an attractive alternative to conventional bearings of sintered-bronze.

In addition to these tests, an investigation is under way on the long-term behaviour of lubricating greases in helical-groove bearings; among the topics examined are the apparent viscosity of greases, their yield-stress values, 'slumpability', creep, evaporation and phase segregation under centrifugal forces. Further theoretical studies into the hydrodynamic behaviour of helical-groove bearings are also under way [4].

G. Remmers

---

[3] 36 000 hours of continuous running (4 years) would represent a life of many hundreds of years in normal domestic use. However, ageing of the grease through exposure to the atmosphere is the limiting factor in practice.

[4] J. Bootsma, The gas-liquid interface and the load capacity of helical-grooved journal bearings, Trans. ASME F (J. Lubrication Technology) **95**, 94-100, 1973; The gas-to-liquid interface of spiral-groove journal bearings and its effect on stability (submitted to the same journal).

G. Remmers is with Philips Research Laboratories, Eindhoven.

The reaction wheels of the Netherlands satellite ANS

J. Crucq

The attitude-control system of the ANS satellite includes a number of actuators. Their purpose is twofold. First, they have to stop the initial spin of the satellite and point the solar panels for the power supply at the Sun. Secondly, they have to point the astronomical instruments on board at the objects to be observed. Actuators for use in a satellite must be small and of low mass; at the same time, however, they have to meet the highest standards of accuracy. A description of the actuators has already been given in an earlier general article on the attitude-control system of the satellite. The article below deals with the reaction wheels in greater detail. They differ from reaction wheels previously used in space vehicles in that a solid lubricant has been used instead of oil or grease.

The mechanical actuators of the attitude-control system for the ANS satellite comprise three reaction wheels [1]. The wheels control the attitude and movement of the satellite by exchanging angular momentum with it, and may be regarded as mechanical 'storage batteries'. They should not therefore be confused with gyroscopes, as used for navigation in ships and aircraft and also in space vehicles. The axis of rotation of each of the three wheels coincides with one of the coordinate axes of the satellite. Each wheel thus forms part of a control system that regulates the movements of the satellite about the corresponding axis. As the three systems are virtually identical, a description of one of them should be sufficient.

After the first, non-recurring operating modes immediately after the launch, the attitude-control system has two main functions: performing the slew manoeuvres of the satellite in scanning for a stellar object and pointing the observation instruments accurately at that object. During the star-pointing mode a number of external disturbance torques — usually very small — have to be compensated (*Table I*).

Table I. Origin and magnitude of the external disturbance torques acting on the ANS satellite during the flight and to be compensated by the reaction wheels.

Aerodynamic forces	$< 10^{-4}$ Nm
Gravitational gradient	$< 2\times10^{-4}$ Nm
Earth's magnetic field	5×10^{-4} Nm
Radiation pressure	negligible

After a discussion of the principle of a reaction wheel the requirements that must be satisfied by the wheels used in the satellite will be given. This will be followed by a discussion of the mechanical and electrical design, with a brief look at the way in which the wheels are incorporated in the attitude-control system and at the results of the life tests on the wheels.

Ir J. Crucq is with Philips Research Laboratories, Eindhoven.

Principle of a reaction wheel

A reaction wheel contains a body with a certain moment of inertia, called the inertia wheel, that can be kept in rotation by an electric motor. Since to every action there is an equal and opposite reaction, the satellite can 'react' against the moment of inertia of the wheel. The resultant torque T is given by the equation of motion for rotations: $T = I\dot{\omega}$, where I is the moment of inertia of the inertia wheel about its axis and $\dot{\omega}$ is the angular acceleration of the wheel.

The attitude of the satellite can be changed by the successive acceleration and slowing down of the wheel (optimal switching or 'bang-bang control', see *fig. 1*). After a change of attitude the satellite returns to the stationary state and the inertia wheel rotates again at the original speed. Frictional forces in the bearings are internal forces in the system, and thus have no effect on this process.

When a disturbance torque T_d acts on the satellite, it follows from the law of the conservation of momentum that:

$$I_{sat}\omega_{sat} + I_{rwl}\omega_{rwl} = \int T_d \, dt,$$

where quantities with the subscript sat relate to the satellite and those with the subscript rwl to the reaction wheel. By appropriately accelerating and slowing down the reaction wheel it is possible to keep ω_{sat} equal to zero in spite of the presence of disturbance torques. If

the compensation of a disturbance torque constantly acting in one direction should cause the reaction wheel to rotate too fast, the wheel can be 'discharged' by means of the magnetic actuator or 'torquer' [2]. The interaction between the torquing coils and the Earth's magnetic field produces a variable torque that slows down the speed of revolution of the wheel. In this way the torquing coils transfer angular momentum to the Earth.

Design specifications

To perform reasonably rapid slew manoeuvres a reaction wheel must be capable of delivering a torque of 10^{-2} Nm (even then a slew of 60° takes about 2 or 3 minutes). On the other hand the torque must be accurately controllable for fine pointing. The smallest torque value that can be generated by a reaction wheel is therefore $\frac{1}{6}\times 10^{-3}$ Nm.

The attitude-control system of ANS works with a sampling frequency of 1 s^{-1}, which means that in every second there is a determination of the torque to be delivered by the reaction wheel during the next second. A 5-bit control command is used, one bit determining the sign of the torque and the other four the magnitude so that for a maximum torque of 10^{-2} Nm the minimum torque is $\frac{2}{3}\times 10^{-3}$ Nm. Since this division is too coarse there is in addition a 'quarter-second mode', in which the desired torque is delivered for only a quarter of a second. Averaged over one second this corresponds to the torque of $\frac{1}{6}\times 10^{-3}$ Nm mentioned earlier. To obtain a torque of sufficient magnitude with such a fine subdivision, the circuit is designed to permit duplication of the most significant bit of the command. For the fine subdivision this gives a maximum torque of $\frac{23}{6}\times 10^{-3}$ Nm per command.

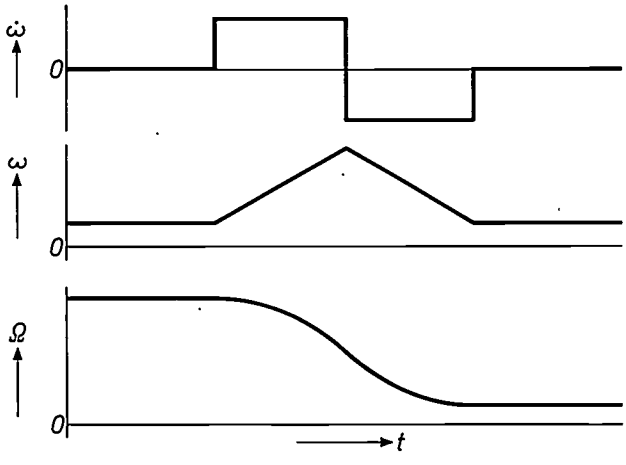


Fig. 1. A change in the angular position Ω of one of the axes of the ANS satellite effected by the appropriate reaction wheel. The upper two graphs show the corresponding variation of the angular acceleration $\dot{\omega}$ and the angular velocity ω of the reaction wheel. Since the moment of inertia of the satellite is about 10^4 times greater than that of the reaction wheel, the movement of the satellite will be 10^4 times slower than that of the wheel.

Computer simulations of the movements to be performed by the satellite have shown that the reaction wheel in the attitude-control system we have adopted must be capable of storing an angular momentum of 0.6 Nms. An important factor here is that the transfer of angular momentum to the Earth with the magnetic torquer, which normally begins at a speed of only 480 rev/min, may sometimes be very poor because of an unfavourable position of the magnetic coil with respect to the Earth's magnetic field. The speed of a wheel can therefore become temporarily fairly high.

The maximum speed of a wheel should not be too high, to ease the load on the bearings and also to limit the gyroscopic forces exerted by the wheel when tilting about its axis, since these forces produce disturbance torques that have to be compensated by the other two wheels. The maximum speed was therefore set at 200 rad/s (about 2000 rev/min). This speed, and the angular momentum of 0.6 Nms mentioned earlier, fix the moment of inertia of the wheel at 3×10^{-3} kgm^2.

An attitude-control system for a satellite using a reaction wheel as an actuator may operate in one of two ways. In the one case there is a direct command to deliver the torque needed to change the attitude, irrespective of the speed, the direction of rotation or the sign of the acceleration of the wheel. In the other case the speed of the wheel is controlled and a torque is obtained by requesting a change in the set value of the speed controller.

The first system is the one we have adopted for ANS. The wheel is driven by a d.c. motor with a permanent-magnet rotor. If the stator current is held constant the torque of such a motor is independent of the speed and is proportional to the stator current. To obtain the desired torque it is therefore necessary to set the stator current to a value corresponding to the torque.

The only complication now is that the frictional torque of the bearings and the hysteresis losses of the motor appear in the attitude-control loop in the form of disturbance torques. If we had chosen the other alternative, the frictional torque would no longer have been a disturbance in the attitude-control system. Instead, however, the speed of the wheel would have had to be accurately controlled over the whole range from 0 to 2000 rev/min, which would have been far from simple.

Apart from the requirements mentioned above there are various 'interface' requirements that a reaction wheel for a satellite will have to satisfy. These include in the first place the volume and weight requirements and compatibility with the digital control system of

[1] P. van Otterloo, Attitude control for the Netherlands astronomical satellite (ANS), Philips tech. Rev. 33, 162-176, 1973 (No. 6).

[2] See page 165 of the article of note [1].

the satellite. The wheels must of course be capable of withstanding the forces and vibrations set up during the launch, and the mechanical operation should not be affected by the temperatures and low pressure ($< 10^{-3}$ Pa or about 10^{-5} torr) prevailing in the satellite during the mission. Finally, and this is no easy requirement to meet, no materials can be used that might cause contamination of the lenses and mirrors of the astronomical instruments on board. This means that only solid lubricants can be used for the moving parts. If this is impossible or undesirable, all oil- or grease-lubricated parts must be contained in a hermetically sealed housing. Materials for purposes such as electrical insulation must be specially proved and reliable plastics, such as PTFE (polytetrafluorethylene, 'Teflon'). Alloys containing cadmium or zinc (such as brass) must on no account be used, because of their high rate of evaporation.

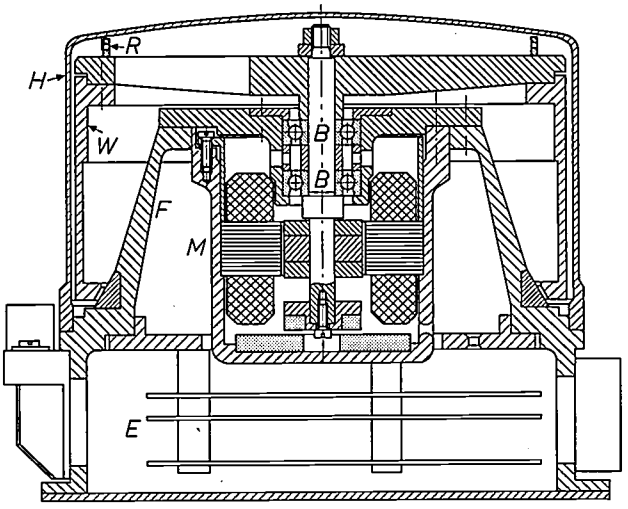


Fig. 2. A cross-section of a reaction wheel. *W* rotating inertial mass, the inertia wheel. *B* bearings. *F* frame housing a brushless d.c. motor *M* and printed-circuit boards *E* carrying the control electronics. The components of the electronic commutator are shown shaded. *H* dust cover. The rim *R* has 240 perforations and is used for accurately measuring the speed and acceleration of the wheel during tests.

Mechanical and electrical design

The general design of the reaction wheel can be seen from the cross-sectional drawing in *fig. 2*. An aluminium frame contains the motor and the printed-circuit boards with the control electronics. The shaft projecting out of the housing carries the inertia wheel, which is completely enclosed by a dust cover. These components, apart from the dust cover, are shown in *fig. 3*.

The rotor consists of a number of magnetized ferroxdure segments. They are fastened together by adhesive to produce a four-pole rotor whose reluctance torque has an amplitude of only 6×10^{-4} Nm ('skewed poles'). The reluctance torque is angle-dependent and has a period of one-eighth of a revolution; the torque has a very small effect at low speeds only. The stator current is commutated electronically. In addition to the advantage of eliminating brush friction and wear, with the associated dust contamination, the electronic commutation has the advantage of delivering signals that can be used to determine the speed of rotation [3]. The circuit for controlling the reaction wheel, details of which will be discussed presently, is mounted on three printed-circuit boards located at the base of the housing.

The inertia wheel consists of a spoked aluminium wheel with a fairly broad stainless-steel rim. The rim is profiled in such a way that the centre of mass of the wheel lies between the bearings, thus minimizing the sensitivity to lateral shock. It was nevertheless found necessary to provide the housing with a plastic buffer strip opposite the lower edge of the wheel, to prevent damage from any tilt of the wheel that might occur during the launch. (The stiffness of the assembly is at its lowest for such movements: the resonant frequency is 160 Hz.) *Fig. 4* shows the complete reaction wheel.

The stiffness of the bearing has to be high, so that the resonances fall at fairly high frequencies (350 Hz for axial movements, 600 Hz for transverse movements), where the vibrational load that can arise during the launch is less than 5 *g*. For the shaft-bearing system we have therefore chosen a pair of angular-contact ball-bearings with a fixed pre-load (*fig. 5*). These bearings are small and very light, readily permitting rotation in either direction at widely different speeds.

To enable the pre-loading of the bearings to be accurately adjusted by means of two spacer bushes of slightly different lengths, the two bearings are mounted on the same side of the motor. Temperature differences in the construction, especially between the two spacer bushes, cause a change in the pre-loading and thus in the bearing friction as well. It is therefore important to keep the temperature differences as small as possible. For this reason as many components as possible are made matt black to give good heat exchange by radiation.

It did not seem desirable to put the reaction wheel in a hermetically sealed enclosure, since this would make the wheel less accessible for measurements before the launching and would in addition involve a weight penalty, quite apart from the sealing problem itself. We therefore decided on dry lubrication for the reaction wheel — the first time it has ever been used for this purpose in a space vehicle.

[3] The design of this motor is due to W. Radziwill and K. W. Steinbusch, who are with Philips Forschungslaboratorium Aachen GmbH, Aachen. The principle of the commutation is discussed by W. Radziwill in Philips tech. Rev. 30, 7, 1969.

Fig. 3. Components of the reaction wheel; from left to right: the motor with bearing, the housing and the inertia wheel.

Fig. 4. The reaction wheel. The whole structure is 150 mm in height and in diameter, and weighs 2795 g. The connector block has an extension cord (the 'pin-saver'), which is used during pre-flight testing to protect the connector pins in the wheel from being damaged by repeated plugging and unplugging.

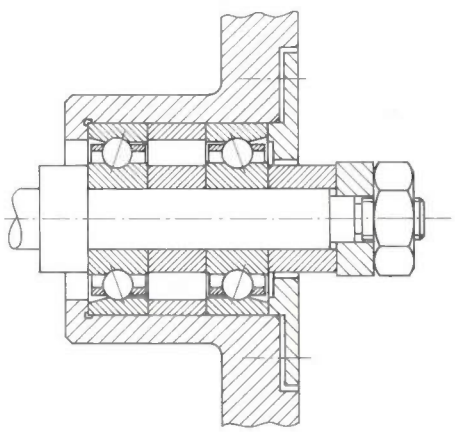


Fig. 5. The shaft bearing of the reaction wheel. The angular-contact bearings are given a fixed pre-load by making the outer spacing bush 3 to 5 μm longer than the inner one. The pre-loading eliminates play in the bearings and gives the whole structure great axial and radial stiffness.

Lubrication of the ball-bearings

Although some general information on the dry lubrication of ball-bearings can be found in the literature, there are no details relating to the application we had in mind for the ANS reaction wheels, in which the bearings are required to carry an inertia wheel with a weight of about 1 kg. The characteristic features of our application are that the wheels are required to operate both in air and in vacuum, and at temperatures varying between −20 °C and +50 °C. The wheels must always be capable of rotating in either direction at varying speeds and accelerations. The required life in air is 1000 hours, followed by at least 2×10^8 revolutions in vacuum. This corresponds to six months of normal operation of a reaction wheel in the ANS satellite.

On the basis of the available data and our own experience with the first series of reaction wheels, it was decided, in consultation with our American advisers, to use a composite of MoS_2 and PTFE, to give the same kind of lubrication as used in the horizon sensor. Problems to be solved for this application were cleaning the ball-bearings, the method of applying the lubricant, running in the bearings and the choice of the pre-load for the angular-contact bearings.

The lubricant was applied by making the bearing cages of 'Duroid' 5813 (70% PTFE, 15% MoS_2, reinforced with 15% fibreglass); the design of the bearings is illustrated in *fig. 6*. To avoid damage to the lubricant film, the pre-load on the bearings is kept relatively low, at a value of 4 N instead of the values between 10 and 30 N used for bearings of this size in normal operation. The frictional torque, which is directly related to the pre-load, therefore remains relatively low ($1\text{-}1.5\times10^{-4}$ Nm).

The friction of the balls against the cage causes the balls and the tracks of the bearing to become gradually coated with a thin film of lubricant. This film reduces the friction in the bearing and prevents cold welding between the balls and the tracks, which could damage the bearing surfaces and cause the bearing to seize up.

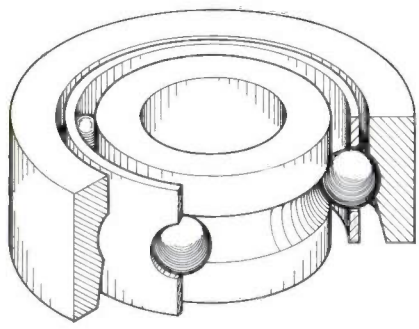


Fig. 6. Diagram of a ball-bearing for the reaction wheel. The cage containing the balls is made from 'Duroid' 5813 (70% PTFE, 15% MoS_2 and 15% fibreglass for reinforcement); the cage is shown as thinner than it is in reality.

The friction in a bearing lubricated in this way is independent of the speed of rotation, unlike an oil-lubricated bearing, in which friction does depend on speed because of the viscous nature of the lubricant. The frictional torque of a well run-in dry-lubricated bearing is only 0.5 to 0.75×10^{-4} Nm, whereas the same bearing lubricated with a small quantity of low-viscosity oil has a frictional torque of 6×10^{-4} Nm. The dry-lubricated bearing does however run rather noisily.

A dry-lubricated bearing of this type will run in air or vacuum. One strange effect that we have noticed, but cannot as yet explain, is the following. When air is admitted to a bearing after it has run satisfactorily for some time in vacuum, the friction initially increases, fluctuating very strongly in amplitude with time, then gradually returning to the original situation. When the bearing is again evacuated, the whole process repeats itself, as can be seen in *fig. 7*.

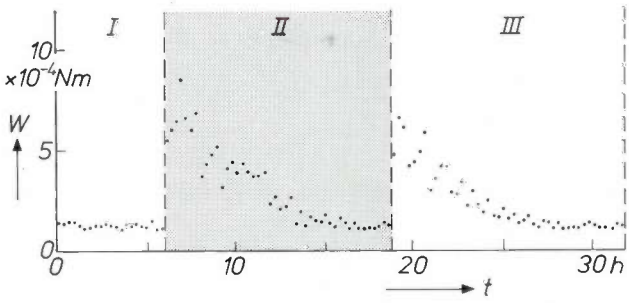


Fig. 7. Bearing friction W as a function of time t. During the intervals *I* and *III* the bearing has run in vacuum, during interval *II* in air.

Electronic control

Control signals for the reaction wheels can come both from the attitude-control logic (ACL) and from the onboard computer (OBC) [1]. The ACL provides a coarse attitude control and can only choose from three torque values (maximum positive, maximum negative and zero). This coarse control always takes priority over the fine attitude control, which is effected through the onboard computer.

We shall now look at the operation of the control electronics with the aid of the block diagram in *fig. 8*. The priority logic *1* determines which of the two incoming commands is to be carried out, and sends to the digital-to-analog converter *2* a signal specifying the magnitude of the torque to be delivered. At the same time it sends to the circuit *4* a signal specifying the sign of the torque. The analog signal for the required magnitude of the torque is compared in the comparator circuit *3* with the average stator current i_{st}. If necessary, switch *5* is then operated, which ensures that the current has the appropriate mean value. The electronic commutator of the motor *6* delivers pulses that report the times of commutation to the circuit *4*, which in its turn controls the phase of the current pulses sent to the motor by switch *5* in accordance with the required sign of the torque to be delivered by the motor.

The commutator pulses also serve to produce a digital speed signal in the circuit *7*. From this the counter *8* forms a digital word once every second, which is sent as housekeeping information to the ground station via the telemetry system of the satellite. The same speed information is passed via the discharge logic *9* to the system that controls the magnetic torquer, which has to reduce the speed of the reaction wheel if it becomes excessive.

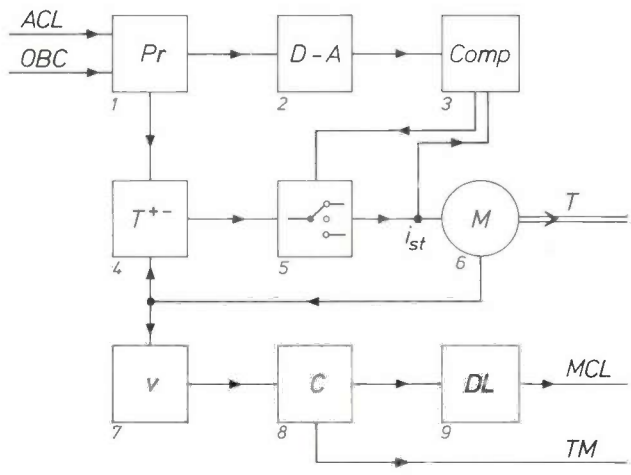


Fig. 8. Block diagram of the control electronics. The input signals may come either from the attitude-control logic (*ACL*) or from the onboard computer (*OBC*). *1* priority logic. *2* digital-to-analog converter. *3* comparator. *4* circuit for controlling the sign of the torque. *5* circuit for controlling the magnitude of the torque. *6* motor. *7* digital speed-signal generator. *8* counter. *9* discharge logic. i_{st} stator current. T generated torque. *MCL* output signal that sets the magnetic torquer in operation. *TM* output signal transmitted via the telemetry system as housekeeping data to the ground station.

The control-logic system consists of integrated circuits, which take very little power; the whole circuit takes no more than 65 mA at a supply voltage of 5 V. The supply voltage for the motor is 20 V; the current required varies between 7 and 122 mA depending on the speed and the torque to be delivered.

Life tests

The useful life of the satellite is determined to a large extent by the life of the reaction wheels, and here the life of the bearing plays the most important part. For a useful life of six months the minimum requirement to be met by the bearing is a total of 2×10^8 revolutions as stated earlier, some in air but mostly in vacuum. The estimated number of reversals in direction of rotation is 6500. The temperature is allowed to vary in the range from —20 to +50 °C.

Life tests on three prototype wheels using dry lubrication with MoS_2 gave good results. It was therefore

finally decided to use this kind of lubrication, although it had never previously been used for reaction wheels. The life tests were concluded after 5×10^8 revolutions ($2\frac{1}{2}$ times the required life).

Life tests are currently being performed with four wheels in the definitive design. By the end of April, 1974, these wheels had completed more than 10^9 revolutions without showing any defect and without the frictional torque rising to a value that could indicate undue wear.

The results of the life tests have given a good impression of the reliability of the construction and the method of lubrication, and all the indications are that this design of wheel will give satisfactory service in the ANS satellite.

Summary. The ANS satellite has three reaction wheels, which are used for changing the attitude of the satellite on commands received from the attitude-control system. Furthermore, to maintain a particular attitude accurately, external disturbance torques can be compensated by the exchange of angular momentum between the satellite and the wheels. A maximum torque of 10^{-2} Nm can be delivered to ensure that slew manoeuvres are performed in a reasonably short time. The minimum torque is $\frac{1}{6}\times 10^{-3}$ Nm. The angular momentum is kept below the maximum value of 0.6 Nms by the magnetic torquer, which reduces the angular momentum of the wheel in good time by interaction with the Earth's magnetic field.

The wheel is driven by a d.c. motor with a permanent-magnet rotor and electronic commutation. The lubricant used is a composite of MoS_2 and PTFE ('Teflon'). This dry lubrication has removed the necessity for hermetic sealing, which is necessary with conventional oil lubrication to prevent contamination of the optical system. The dry lubrication works well both in air and in vacuum, and at temperatures varying between —20 and +50 °C. In life tests on four prototype models a life of at least 10^9 revolutions has been recorded, which is more than five times the required life.

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.

J. Bloem: Electrical properties of flux-grown rutile (TiO_2) crystals.
Philips Res. Repts. **28**, 596-604, 1973 (No. 6). *E*

P. M. Boers: Comment on determination of the velocity field characteristic for *n*-type indium phosphide from dipole-domain measurements.
Electronics Letters **9**, 134-135, 1973 (No. 6). *E*

J. van den Boomgaard: Preparation and some properties of monodisperse two-phase in situ composites from quaternary melts in the Fe-Co-Cr-C system.
Philips Res. Repts. **28**, 605-617, 1973 (No. 6). *E*

E. Bruninx: The ^{85}Kr leak test: An improved detection method.
Int. J. appl. Rad. Isot. **24**, 359-360, 1973 (No. 6). *E*

J. Cornet & D. Rossier: Properties and structure of As-Te glasses, I. Glass-forming ability and related properties, II. Local order parameters and structural model.
J. non-cryst. Solids **12**, 61-84 & 85-99, 1973 (No. 1). *L*

J. P. Deschamps & A. Thayse: Applications of discrete functions, Part I. Transient analysis of combinational networks.
Philips Res. Repts. **28**, 497-529, 1973 (No. 6). *B*

I. Flinn: Piezoelectric ceramics.
Electron (GB) No. 32, pp. 59 & 62-64, 12 July 1973. *M*

M. J. C. van Gemert: High-frequency time-domain methods in dielectric spectroscopy.
Philips Res. Repts. **28**, 530-572, 1973 (No. 6). *E*

K. H. Härdtl & D. Hennings: Wechselwirkungen zwischen Gefüge und Gitterstruktur in der ferroelektrischen Mischkristallreihe $PbTiO_3$-$PbZrO_3$.
Science of Ceramics **6**, VII/1-15, 1973. *A*

J. A. Kerr, J. A. G. Slatter & D. Vinton: An f_T anomaly.
Electronics Letters **9**, 338-339, 1973 (No. 15). *M*

D. J. Kroon: The national air pollution monitoring network in the Netherlands.
La Chimica e l'Industria **55**, 49-52, 1973 (No. 1). *E*

G. Le Floch & H. Arnould: Electroluminescence dans une hétérojonction ZnTe-ZnSe.
Solid-State Electronics **16**, 941-944, 1973 (No. 8). *L*

A. Milch: On the formation and thermal stability of Bi_2O_3 films.
Thin Solid Films **17**, 231-236, 1973 (No. 2). *N*

J. G. J. Peelen: Relation between microstructure and optical properties of polycrystalline alumina.
Science of Ceramics **6**, XVII/1-13, 1973. *E*

M. J. Sparnaay, A. J. van Bommel & A. van Tooren: Auger electron spectroscopy as a tool for measuring the diffusion of foreign atoms in solids near their surface.
Surface Sci. **39**, 251-254, 1973 (No. 1). *E*

W. Tolksdorf & F. Welz: Über die Züchtung von galliumsubstituierten Yttrium-Eisen-Granat-Einkristallen aus schmelzflüssiger Lösung bei konstanter Temperatur.
J. Crystal Growth **20**, 47-52, 1973 (No. 1). *H*

T. S. te Velde & J. Dieleman: Photovoltaic efficiencies of copper-sulphide phases in the topotaxial heterojunction copper-sulphide — cadmium-sulphide.
Philips Res. Repts. **28**, 573-595, 1973 (No. 6). *E*

J. A. Weaver: A research worker's view on the future of automatic reading machines.
AGARD Conf. Preprint No. 136, 13/1-8, 1973. *M*

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 5/6

Monitoring the quality of surface water

D. J. Kroon and M. Q. Mengarelli

'... and all the waters that were in the river were turned to blood.
And the fish that was in the river died; and the river stank, and the Egyptians could not drink of the water of the river...',
from Exodus 7:20 and 21, describing the first of the ten plagues of Egypt.

Introduction

In recent years we have become aware of the need in our industrialized society for economy in the consumption of certain materials and intelligent management of their resources. One of these materials, which is vitally important to human welfare and calls for careful management, is surface water, that is to say the water of rivers, lakes, etc. Not only is this water consumed in large quantities, but it is also the natural environment for fish and other creatures — some species of which are used as food — and is part of man's recreational environment.

Surface water is used far more than formerly for all

Dr D. J. Kroon is with Philips Research Laboratories, Eindhoven. Ir M. Q. Mengarelli is with the Italian Philips organization at Monza, Italy.

Title picture: Alarmed fishermen show the Pharaoh dead fish from the Nile. Engraving by Caspar Luyken (1672-1708). (Rijksprentenkabinet, Amsterdam.)

kinds of purposes — for the preparation of drinking water, for irrigation in agriculture and horticulture, for cooling and for industrial processes. Formerly the water needed could be obtained from wells and ground water, but the quantities required have gradually become so great that underground water reserves are no longer sufficient. What is more, increased water consumption also entails greater amounts of industrial and domestic effluents (*fig. 1*). Since all the waste water is not at present purified sufficiently for it to be discharged harmlessly, the pollution of surface water — more and more of which is being used as a source of drinking water — may be increasing. This implies that the authorities responsible for the control of water resources are in ever greater need of information concerning the quality of the water. One of the methods by means of which this information can quickly be obtained is to use a network of automatic monitoring stations. Networks of this type are already operating in some parts of the world, such as the rivers around the city of Tokyo.

In this article we shall attempt to outline the problems involved in controlling the quality of surface water, with special reference to the role that monitoring networks can play in this respect, to the manner in which they can best be installed, and to the requirements to be met by the measuring equipment used in them.

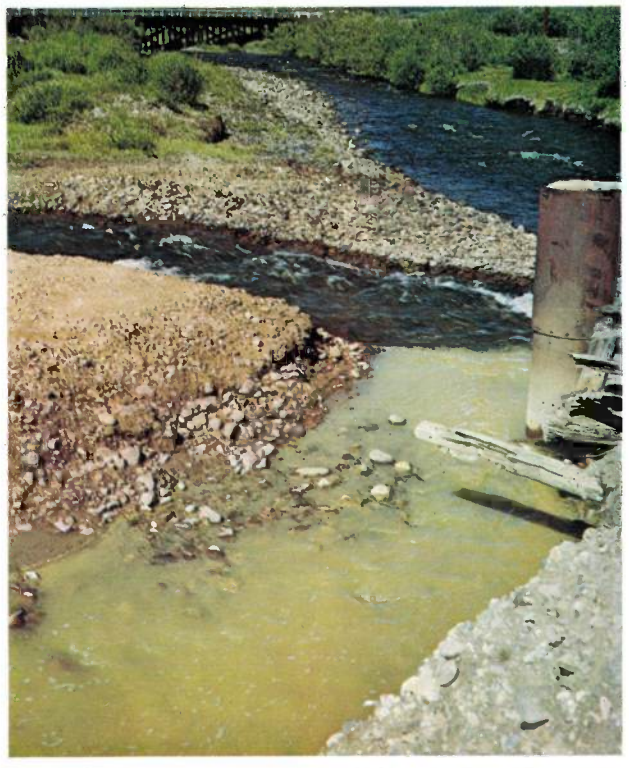


Fig. 1. Pollution caused by industrial effluents (from the mining industry in Colorado, U.S.A.).

Purpose and function of a monitoring network

The main purpose of a monitoring network is to provide the results of measurements in a form that permits a decision to be made on whether or not to warn the authorities responsible for controlling the quality of the water. The warning may be a long-term indication or it may be a short-term alert. A long-term warning is given when the concentration of certain substances shows a gradually increasing trend, whereas a short-term alert is issued when the concentrations of certain substances suddenly rise to a value higher than that expected within the 'normal' variations. This may be the result of the unexpected and prohibited dumping of waste products, of some accident in an industrial process or in a treatment plant for waste water, or during the transportation of toxic substances. In such an event immediate measures must be taken to limit the damage to the environment. These measures may include the closing of neighbouring watercourses so as to confine the pollution to a small area, shutting off the supply of surface water to drinking-water reservoirs, flushing the polluted watercourses, warning industrial users and farmers that the water quality is sub-standard, and possibly tracing the polluter. A monitoring network makes it possible not only to record the water quality but also to take swift corrective action on the basis of the measurements.

The discovery of a slow change in the measured values can also lead to an improvement in quality, since the effect of the corrective measures adopted can be derived from these measurements. At the same time they make it possible to determine the places where improvement is necessary as well as the places where waste water can permissibly be discharged. The natural and artificial impurity content of the water can be ascertained and water quality standards established. In this way a monitoring network can help to keep up the quality of the fresh water in both the short and long term.

Information required from the monitoring network

Before the quality of water can be judged, a very large number of parameters need to be known. What is meant by water of good quality depends of course on the purpose for which the water is to be used, and also on whether the water is going to be used immediately. In the case of drinking water, for example, it is necessary to be certain that the water contains no toxic substances. The same requirement applies to water for recreational purposes and to irrigation water for agriculture and horticulture. If the water is to be stored,

e.g. in impounding reservoirs or lakes, it is vital that there should be little risk of eutrophication — the drastic growth of certain algae ('bloom') due to the presence of relatively high concentrations of nutrients. As the algae die off, decomposition can set in, which depletes the water of oxygen, making animal life in the water impossible and producing malodorous gases.

When all these factors are taken into account it is possible to select various groups of parameters that can be used for evaluating the quality of the water and which it should be possible to determine. These comprise a number of simple physical and chemical parameters, parameters relating to oxygen balance, eutrophication parameters, and also the concentrations of the malodorous components, mineral oils and phenols, certain trace elements and organic micropollutants.

The *physical parameters* involved are first of all water level, flow rate and temperature, i.e. standard hydrological data. Various other parameters of importance include the solar radiation and the rainfall in the catchment areas. Solar radiation is an important factor in the photosynthesis that takes place in the water in algae and other water plants. When combined with data relating to dissolved oxygen and oxygen demand the measurement of solar radiation provides information on the oxygen dynamics in natural water. The measurement of rainfall is important for predicting not only water levels and flow rates but also the concentrations of certain substances that are drained into the water from agricultural land, such as phosphates, nitrates and pesticides. After heavy rainfall the turbidity of the water may also change considerably, owing to more sand and clay entering the river and to sludge from the bottom being stirred up by the faster flow rate. This sedimental sludge can adsorb relatively large quantities of metals and pesticides.

The *simple chemical parameters* are acidity (*p*H), total ion concentration (conductivity) and the content of specific ions that may be present in abnormal concentrations (e.g. Cl^-, CN^- and F^-).

The oxygen content of surface water is determined by the extent to which processes that supply oxygen and those that consume oxygen compensate one another. The oxygenating processes include aeration and photosynthesis. Oxygen is consumed by the metabolism of higher plant and animal life, and is also consumed as a result of biodegradation, i.e. the biological breakdown of organic matter by micro-organisms. This form of oxygen consumption is of great practical interest, since it is possible to control it by regulating the quantities of organic matter discharged.

The oxygen demand of water can be measured by the natural process (Biochemical Oxygen Demand or BOD). If the determination is made by adding chemical reagents the COD value is found (Chemical Oxygen Demand) [1].

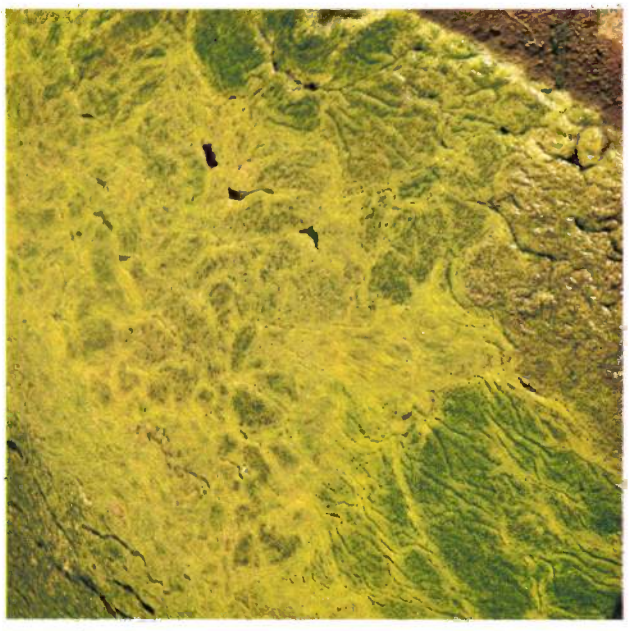


Fig. 2. Proliferation of algae caused by eutrophication (excessive nutrients) in water. When these algae die off, the oxygen in the water is used up and the result is 'dead' water.

Eutrophication — briefly mentioned above — is the proliferation of certain algae caused by an excess of 'nutrients' in the water, for example phosphates, nitrates and carbonates (*fig. 2*). When these algae die off, large amounts of oxygen are used up in the breakdown of the organic matter and this can make the oxygen content fall to almost zero. Animal life in the water is then no longer possible (the water is 'dead') and organic matter is broken down not by oxidation but by other means, which can give malodorous results. It is useful to determine eutrophication parameters if the water has to be stored virtually without movement over a longish period in an impounding reservoir or lake, a situation conducive to eutrophication.

The content of *malodorous compounds*, or of substances that can readily be converted into malodorous compounds (amines), is important at places where water is admitted to recreational lakes, drinking-water reservoirs, etc.

Rapid measurements of *mineral oils* are necessary to trace prohibited discharges of oil and to confine the oil slick to a small area. It is particularly important to prevent valuable nature reserves from being spoilt by oil slicks. Apart from floating oil, oil in emulsified and dissolved form should also be measured. Analysis of the oil to determine its composition, sulphur content and trace elements (such as vanadium) often makes it possible to establish the origin of the oil discharge.

[1] See the article by P. F. Butzelaar and L. P. J. Hoogeveen in this issue, page 123.

Phenols are important in surface water used for the preparation of drinking water because the untreated water is sterilized with chlorine. Any phenols present will then be chlorinated to form chlorophenols, which are notorious for the unpleasant taste they give to drinking water. Even concentrations as small as a few micrograms per litre may affect the taste of the water.

Certain *trace elements* and *organic micropollutants* have to be removed from surface waters because they are poisonous. Some of these compounds appear in the two lists compiled by the International Rhine Commission. The 'black' list contains substances that must never be allowed to get into surface water, such as mercury, cadmium, organic chlorine compounds and carcinogenic substances. The 'grey' list contains substances that should be kept out of surface water as far as possible. These include the phenols as mentioned earlier, substances that affect taste and smell, mineral oils and a large number of metals, such as zinc, copper, nickel, chromium and lead.

Designing a monitoring system

The foregoing makes it clear that the complete quality control of surface water requires detailed knowledge of the occurrence of a large number of substances. This knowledge can of course only be obtained from measurements made at a limited number of points and times, chosen as far as possible in such a way as to gather the maximum of relevant information for the minimum of expense and effort. In setting up a monitoring system it is therefore necessary to take into account the topographical situation, and to determine the requirements to be met by the detectors for time resolution, accuracy and detection limit. Finally the design must make proper allowance for the availability of good and reliable methods of measuring the quantities on which information is required.

Basically there are two forms of measurement possible in a monitoring system: manual and automatic. In the first case water samples are taken from time to time at various places and sent for analysis to a central laboratory; in the second case, water samples are analysed on the site and the result of the analysis is reported to a central point by a telemetering system. In such an automatic system unattended monitoring stations are used.

Manual sampling

Manual sampling is of course the oldest method, and is still used in many parts of the world. For a number of analyses that are difficult to automate it is also at present the best method. The water samples are collected in glass or plastic bottles and sent for analysis, after the addition of conserving reagents if necessary. The installation of a monitoring network in which sampling is done manually obviously calls for a well equipped analytical laboratory, which is an expensive undertaking. On the other hand, the installation at the sites where the samples are taken is inexpensive or may even cost nothing at all. Advantages of manual sampling are that all the analytical techniques that are known can be used and that the system is readily adaptable with regard to the number of components and the number of monitoring sites. It is very easy to adapt the analysis method to the requirements of the moment, to raise or lower the sampling frequency, or to take the samples at different sites.

These advantages are offset by certain disadvantages. In the first place the sampling frequency is low. Sometimes it may not be possible to take even one water sample a day at a given point. This situation can be improved by using mechanical sampling devices, which take a sample at fixed times and preserve it under defined conditions. We shall return later to the use of this mechanized sampling equipment. A second disadvantage is the very long time lag between sampling and obtaining the result of the analysis. Continuous round-the-clock monitoring is therefore not easy to arrange. Because of the low sampling frequency and the time lag, no correlation can be determined between water compositions at different places, which makes it difficult to trace polluters. Finally, in dense or very extensive networks many people have to be employed to handle the very large number of samples, which makes the running costs very high.

It is obvious that a monitoring system with manual sampling is not a suitable system for signalling and reporting sudden changes in water quality. These systems are, however, suitable for the measurement of slow variations (trends). A manual system should also be used before an automatic system is installed. The manual measurements will indicate which components need to be continuously and automatically measured, what the lowest and highest concentrations are, and it can also provide information about the best sites for the monitoring stations.

Automatic systems

In automatic systems the sample is analysed on site and the result is transmitted by telephone, telex or radio to the central agency. The measuring instruments are directly connected to the telemetering system, and the data are collected without human intervention by a central computer. This calculates the deviations from the 'normal' situation and passes them on to the operating personnel.

An automatic system offers various facilities and

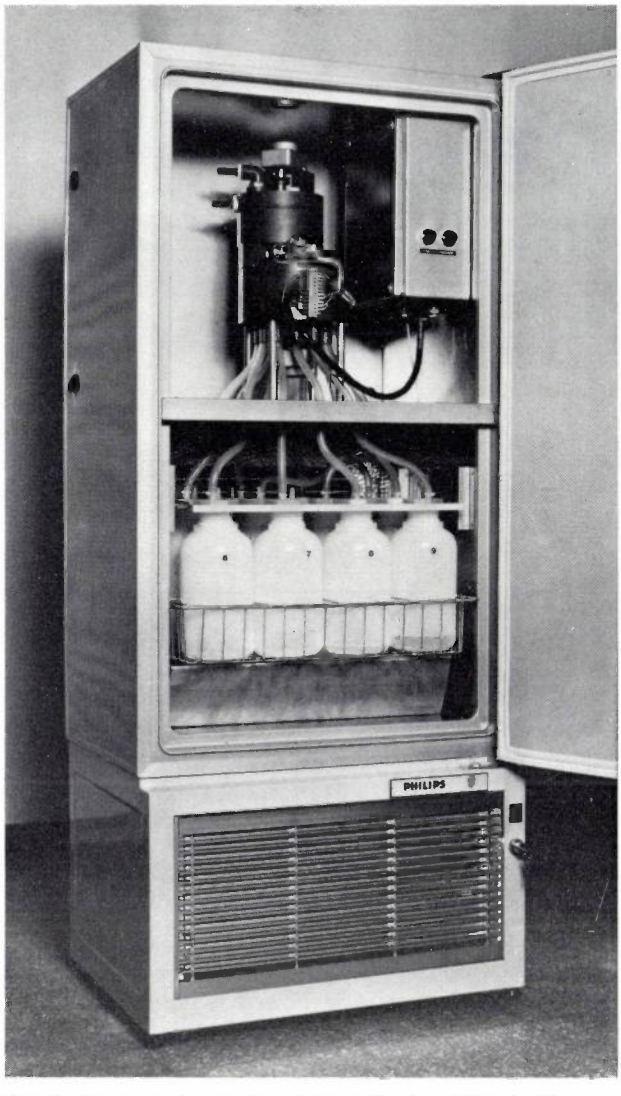


Fig. 3. Mechanical sampling device. Twelve 2-litre bottles are contained in a refrigerated space. At intervals preset by a timing mechanism, one of the bottles is filled with a water sample. This device can also be built into an automatic water-monitoring station. If there are deviations in the water quality, the central control room can send an instruction to the device to store a water sample.

advantages which a manual system cannot provide. In the first place it gives an instantaneous survey over a large area of water flows. In the event of sudden and serious pollution, due for example to an accident, the system makes it possible to compute the optimum set of measures needed and to carry them into effect. In areas of complicated hydrographic structure this can be a particularly difficult and time-consuming problem without a computer, and mistakes can easily be made.

A second advantage is that, because of its high sampling frequency, an automatic system makes it possible to compute time correlations. This facilitates the tracing of clandestine polluters.

To perform calculations of the type referred to, the computer has to be programmed with a mathematical model of the entire hydrographic area. A large area will as a rule be divided into smaller units of known input and output, for each of which a separate model is drawn up. In one of the following sections we shall give an example of a typical model.

The running costs of an automatic system are low, especially if unmanned stations are used, connected to a central computer programmed to carry out many of the preparatory calculations. One operator is then sufficient to supervise and control a large network. (The national air-pollution monitoring network in the Netherlands is an example for comparison; this has 250 monitoring stations spread over an area of 30 000 km^2 and operates completely automatically [2].)

A difficulty encountered with a monitoring network using unmanned stations is that the requirements to be met by the analytical equipment for reliability and specificity are much more difficult to meet. This means that special instruments have to be developed for certain measurements. Another disadvantage is of course the system's lack of flexibility. It is not easy to add components to the list of pollutants to be determined, and changing the site of a monitoring station generally means that electricity and telephone cables have to be relaid. Before an automatic network is built careful studies must therefore be made to decide which quantities are to be measured and where the monitoring stations can best be sited.

The best system will be of mixed type, partly automatic and partly using manual sampling. A mixed system can offer reasonable flexibility without being too expensive, and is capable of supplying fast and reliable information about the quality of the water. Not all samples analysed manually need be collected manually. An automatic monitoring station can be equipped with a mechanical sampler that takes a water sample and stores it on receipt of an instruction from the operator (*fig. 3*). Such samples may if necessary be taken to serve as evidence in legal proceedings against clandestine polluters.

Types of station

The stations in an automatic system for the monitoring of water quality do not all have to be of the same type. To obtain rapid and accurate information about the water at discharge sites it will be necessary to situate monitoring stations near the known pollution sources, but it may be sufficient to measure only a few components there.

Simple monitoring stations may also be sufficient at other specific points in the area where the water quality

[2] See for example Philips tech. Rev. 33, 194, 1973 (No. 7).

is being monitored, for example at the boundaries of the model areas referred to above. At each of these units the 'input' and 'output' must be known, both for the quantity of water itself and for the number of substances present in the water. Just which parameters can best be determined at the boundary of two regions depends of course on the expected discharges in the upstream area and on the use that may be made of the water in one of the next areas.

These monitoring sites, which contain special equipment and are located at special sites, are referred to as *key points*. In addition, complete monitoring of water quality requires a group of monitoring sites in the network where a large number of parameters are measured periodically; these are known as *main points*. Complete monitoring of this nature is only possible in a network that contains monitoring points of *both* types, all of which are connected on-line to a central computer that not only stores the data but also uses a mathematical model of each unit area to calculate what is to be expected in the next sub-area.

In addition to monitoring stations of the two types just mentioned, which are continuously and directly involved in monitoring the water quality, stations of a third type are needed, referred to as *reference points*. These must be capable of measuring *all* the relevant parameters. They differ from the two other types of station primarily in the different use made of the data they supply. The data are used for periodically checking whether the information obtained from the other monitoring stations still gives a good overall picture of the quality of the water in the monitored area, in other words, whether the most relevant parameters are still being measured. Since reference-point stations are used for a different purpose, the result of the measurements can be made available at a much lower frequency than from the other monitoring stations.

Example of a mathematical model of a flowing river

As an example of a mathematical river model we shall outline the relation that should exist between the BOD value and the content of dissolved oxygen in a particular segment of a river when the degradation of organic matter takes place normally.

When the oxygen concentration is greater than zero, the change in the BOD value is described by the differential equation

$$\mathrm{d}\beta/\mathrm{d}t = -k_1\beta, \qquad (1)$$

where β is the BOD value and k_1 is a reaction constant dependent on the temperature T:

$$k_1 = A \exp(BT),$$

where A and B are constants. If the concentration of dissolved oxygen is zero — in other words if the water is completely anaerobic — a slow breakdown takes place by other means than the consumption of oxygen, so that the BOD value does show a slight decrease. In our example, however, we put $\mathrm{d}\beta/\mathrm{d}t = 0$.

The concentration of the dissolved oxygen depends on many factors, first of all on temperature. The saturation concentration c_s is given approximately by:

$$c_s = 4000 \exp(-0.02\,T) \text{ mg/l}.$$

The oxygen concentration in the water may *increase* because of the solution of atmospheric oxygen in the water (aeration) and the oxygenation resulting from photosynthesis by algae and water plants during the day. The increase due to aeration is proportional to the difference between the saturation concentration c_s and the prevailing concentration c, and is given by

$$k_2(c_s - c) \text{ mg/l s}.$$

Apart from c_s, the value of k_2 in this equation is also temperature-dependent. We describe the average contribution of photosynthesis with a term P, which depends only on the intensity of the sunlight and on the temperature.

The oxygen concentration in the water decreases because oxygen is consumed in the breakdown of organic matter by micro-organisms, and is also consumed by higher organisms. The oxygen consumed by higher organisms can simply be represented by a term M, which depends only on temperature. The decrease of oxygen concentration due to the breakdown of organic matter is of course equal to the BOD decrement $-k_1\beta$ when $c > 0$, and is equal to zero when $c = 0$.

Summarizing, the change in the oxygen concentration as a function of time is therefore given by:

$$\mathrm{d}c/\mathrm{d}t = k_2(c_s - c) - k_1\beta + P - M. \qquad (2)$$

Combination of (2) with the solution of (1) gives:

$$\mathrm{d}c/\mathrm{d}t = k_2(c_s - c) - k_1\beta_0 \exp(-k_1 t) + P - M,$$

where β_0 represents the BOD value at the time zero.

When a few simplifying assumptions are made, this differential equation can easily be solved. We assume first that the temperature of the water is constant (i.e. that no warm cooling water is discharged). We assume further that the intensity of the sunlight is constant, which implies constant photosynthesis, and that no organic material is discharged into the water. The solution is then:

$$c = \{c_s + (P - M)/k_2\}\{1 - \exp(-k_2 t)\} + \\ + \{k_1\beta_0/(k_2 - k_1)\}\{\exp(-k_2 t) - \exp(-k_1 t)\} + \\ + c_0 \exp(-k_2 t),$$

where c_0 is the oxygen concentration at the time zero.

The variation of the oxygen concentration as a function of time is now calculated for a number of values of β_0 (the BOD value at the beginning of the river segment) and for two values of the oxygen concentration c_0 at this same site (i.e. for one very low value, 2 mg/l, and for one reasonably high value, 8 mg/l). The water temperature was assumed to be 15 °C, the reaction constants k_1 and k_2 were taken to be 0.142 and 0.62 per day respectively, and an average value of 2.76 mg/l per day was taken for $P - M$ [3].

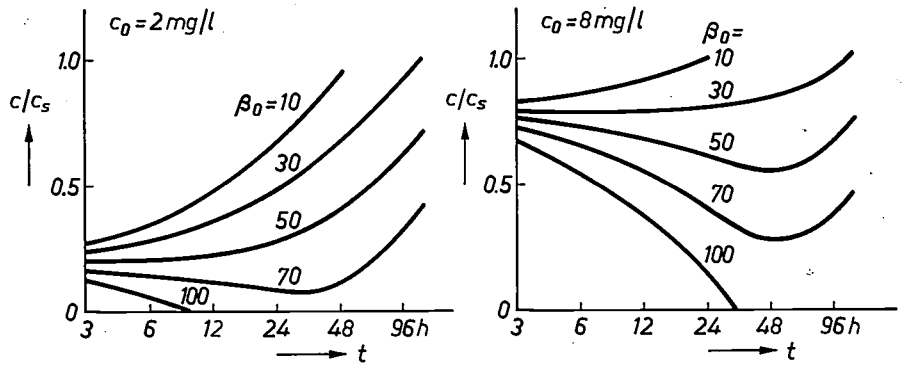


Fig. 4. Theoretical variation of the oxygen content in surface water as a function of time when the natural processes take place without interference. The oxygen content c is plotted as a fraction of the maximum value c_s that can occur at the prevailing temperature (here 10 mg O_2/l). The curves were calculated for various values of the BOD value at the time zero (β_0). The figure on the left shows the situation where the oxygen concentration at the time zero is 2 mg/l (20% of the saturation value); for the right-hand figure the value is 8 mg/l (80%). At β_0 values of 100 and more the water becomes anaerobic. The rapid increase in the oxygen content in the left-hand figure is due to increased aeration.

These calculations resulted in the curves given in *fig. 4*, which thus present a schematic picture of the unperturbed variation of the oxygen concentration with time.

Using a mathematical model of this type the central processor of an automatic surface-water monitoring system is able to carry out computations on the basis of the data received from all monitoring stations. When supplied with data relating to heating or cooling, the breakdown of organic matter, reaction constants for this breakdown, and data relating to the take-up of atmospheric oxygen derived from data obtained through one or more upstream monitoring stations, the computer can use equation (2) to predict the state in which the water will pass the downstream monitoring station some time later. If this prediction indicates a danger to living organisms in the river, an alert can be sent out to stop discharges of organic material or cooling water.

If a comparison of the predicted values with the measured values reveals statistically significant deviations in the oxygen concentration, BOD or calculated reaction constants, something unusual must have happened. This may be an unexpected discharge of organic matter somewhere between the two monitoring stations, but it may also arise from poisoning of the biological breakdown mechanism, for example by minute traces of an agricultural pesticide. In any event, such deviations call for action.

The above example of a mathematical river model is of course greatly simplified. The computer models being worked on at present are in general much more complicated, both in the number of parameters and in the interactions between them. The example given is no more than an illustration.

Detectors for an automatic monitoring station

The detectors installed in an automatic monitoring station for determining water composition are mostly based on the same methods as those used in an analytical laboratory. However, not every analytical technique used in a laboratory can be used in an automatic monitoring system. Thus, although the detectors are based on the same principles as the laboratory instruments, the actual arrangement will often be quite different. The reason for this difference lies in the way the analytical problem is approached. In the analytical laboratory the preference is for standard instrumentation, such as spectrophotometers, X-ray fluorescence equipment and gas chromatographs, and the sample to be tested is modified to match the equipment. Extraction may be required in order to remove interfering components and concentrate the component for analysis, colouring reagents may have to be added, the *p*H may have to be adjusted to the required value, and so on. The modified sample is then in a form in which it can be analysed by the available techniques.

[3] B. Davidson and R. W. Bradshaw, Environ. Sci. Technol. 1, 618, 1967.

Fig. 5. Automatic monitoring station on the river Lambro near Milan.

In automatic analytical equipment used in unmanned monitoring stations the sample cannot be pretreated, or can only be pretreated to a limited extent. Moreover, since one specific determination has to be repeated time and time again, it is not the sample that has to be adapted to fit the instrument, but the instrument to the sample. A consequence of this is that the determinations have to be performed against a high and strongly fluctuating background. The marked variation in the concentrations of components that can interfere with the determination is one of the great problems in the development of analytical equipment for automatic water monitoring stations. The only analytical techniques that can be used are therefore those that have a 'built-in' selectivity.

In some cases it may be useful to base a detector on a principle not usually used in the analytical laboratory. Analytical instruments that give reliable results for a long time under laboratory conditions may often break down after a short period of operation in a 'natural' environment. Such instruments can usually be made to operate reliably by means of ultrasonic cleaning and frequent calibration (which must also be done automatically), but it is better to look for completely different methods that are specific, fast and sensitive. An example of a detector developed to operate for a long period completely without supervision is the COD monitor described in the following article in this issue [1].

An important general problem is the method of sampling. How can a representative sample be obtained from a large river, how is a sample to be transported, stored and processed, should the sample contain suspended particles in the water or should it be filtered before analysis? This question is particularly important for the determination of heavy metals and pesticides, much of which is adsorbed by the suspended matter. Opinions tend to differ on the answers to questions relating to the appropriate methods of sampling.

Finally a word about the speed, detection limit and the maintenance requirements of the monitors.

If an automatic network were to be equipped with monitors that only presented the results of a measurement after a considerable time lag, there would not be very much advantage in having an automatic monitoring system: the faster a determination is made and the result transmitted, the earlier it is possible to take action in the event of trouble. For the determination of oil it may even be desirable to have a sampling frequency of four times an hour, in which case the monitor must be capable of presenting a result for each measurement within 15 minutes. Heavy metals, such as mercury and cadmium, should possibly be determined at a frequency of once an hour.

Some monitors are required to have an extremely low detection limit since it is necessary to be able to determine very low concentrations of certain pollutants, such as heavy metals and pesticides. The detection limit for mercury, for example, should be lower than 0.1 μg/l, and for certain pesticides even lower than 0.01 μg/l. Even with current manual methods it is not easy to determine such low concentrations of these substances.

Like the monitors used in automatic systems for air-pollution monitoring, those used in water-monitoring systems must be capable of operating unattended for a considerable time. They must of course also provide a signal that can be transmitted by a telemetering system, and remote automatic calibration must be possible. In the water-monitoring stations this is done by means of calibrating liquids that can be passed through a particular monitor to check its operation on receipt of an instruction from the central authority.

Example of an automatic monitoring station

A water-monitoring station suitable for operation in an automatic network as described above has recently been developed by the Philips Scientific and Analytical Instruments Group (*fig. 5*). At present only a limited number of chemical and physical parameters can be monitored in this station, but it will in future be extended to undertake all the required monitoring functions.

Fig. 6 shows a schematic cross-section of the monitoring station. The water is pumped out of the river by a floating pump *P* and flows through a measurement

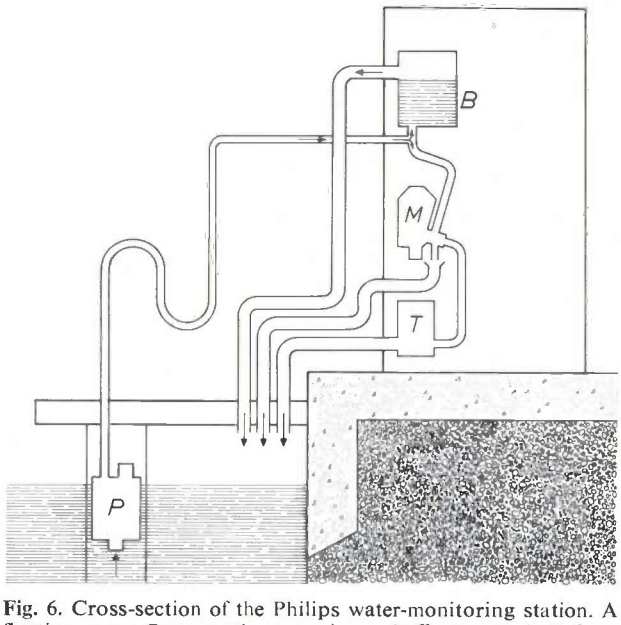


Fig. 6. Cross-section of the Philips water-monitoring station. A floating pump *P* pumps the water into a buffer reservoir *B*, from which the water then flows through the metering block *M* and the turbidity meter *T*.

unit *M* (*fig. 7*). This contains electrodes for measuring the *p*H, the oxygen concentration, the oxidation-reduction (redox) potential and the concentrations of certain ions (e.g. Cl^-) as well as a thermometer for measuring the temperature of the water. In addition to the measurement unit there is a measuring cell for determining the conductivity (on the right in the photograph) and below it a turbidity meter, whose operation is based on the measurement of light scatter. All measured values can both be recorded on site and transmitted to a remote control room.

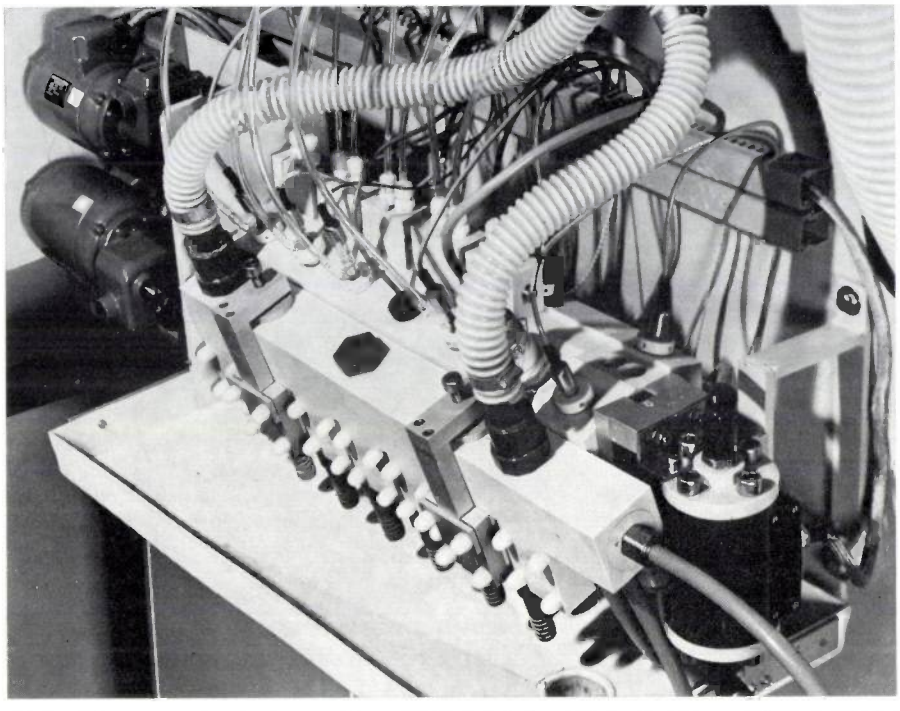


Fig. 7. Measurement unit of the water-monitoring station. The water is pumped through the thick pipes. Behind the pipes can be seen the heads of the measuring electrodes and of the reference electrodes. The thin tubes are for the calibrating fluid feed. In the right foreground is the conductivity-measuring cell.

The station also has facilities for the automatic calibration of all sensors. This is done by means of calibration liquids contained in bottles in the upper part of the compartment. Automatic ultrasonic cleaning of the electrodes takes place hourly; if this were not done they would very soon become covered with slime and algae, and give unreliable readings. On receipt of an instruction from the central monitoring room a water sample can be placed in a refrigerator, and collected later for analysis in a laboratory. This allows components that are not yet automatically measured in the system to be determined, and it also enables a regular check to be kept on the overall functioning of the station. Sensors can also be installed in the station for measuring such quantities as wind direction and speed, water level, flow rate and solar radiation.

The monitoring station has been tested for many months at various places, including sites along the river Lambro (near Milan), where the station was found to operate well for a period of four to five weeks without maintenance.

Summary. A system for monitoring the quality of surface water should provide prompt information about a number of simple physical and chemical parameters, parameters relating to oxygen balance and eutrophication, and also to the concentrations of malodorous components, mineral oils, phenols, certain trace elements and organic micro pollutants. A combination of manual and automatic measurement appears to be the best. Many of the monitoring stations in an automatic network only measure one or two parameters, while a smaller number measure a large number of parameters. The first type should be situated at such places as discharge sites and at the boundaries of the monitored area for which a particular mathematical model has been drawn up. The central computer is programmed with this model, enabling it to compute whether a dangerous situation may arise further downstream. Apart from these two types of monitoring station, which are directly involved in monitoring the quality of the water, a third type is needed for checking whether the data obtained really give a good picture of the water quality. A description is given of a water-monitoring station developed by Philips. The measuring instruments in a monitoring station must be suitable for measuring samples that fluctuate considerably in composition.

A new method of measuring the oxygen demand of water

P. F. Butzelaar and L. P. J. Hoogeveen

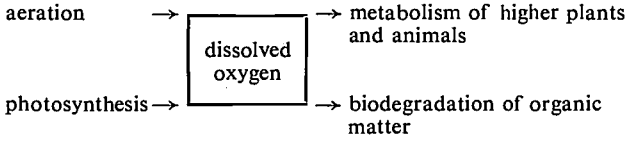
The article below describes a fast and simple method of measuring automatically how much oxygen is needed for the complete oxidation of organic matter present in surface water. This quantity is a measure of the amount of oxygen that should be added to the water to compensate for the consumption of oxygen in biodegradation.

Introduction

To monitor the quality of surface water it is necessary to know certain properties of the water. Because of the important role which oxygen plays in so many biological processes, attention has been devoted first and foremost to the parameters affecting the content of dissolved oxygen [1]. At the present time there is also a growing interest in the content of toxic substances, in particular heavy metals (such as mercury, cadmium, copper, lead and zinc) and pesticides containing chlorine (such as DDT, aldrin and polychlorobiphenylene).

Although most attention has been paid over the years to the content and consumption of oxygen in surface water, the methods of measurement are not all equally suited to the purpose. In this article we describe a new method of measuring the quantity of oxygen used up in surface water through the breakdown (biodegradation) by micro-organisms of organic matter present in the water. The water will only retain its quality if an equal or greater quantity of oxygen is supplied.

The diagram below gives a simplified representation of the oxygen-giving and oxygen-using processes that determine the quantity of dissolved oxygen in the water.



Aeration, photosynthesis and metabolism are processes that are not in principle very amenable to control. The biodegradation of organic matter — that is to say the self-purifying capacity of water — depends to a very great extent, however, on the amount of organic matter present in the water, originating for example from the discharge of organic waste matter. If the concentration of such substances is excessive the dissolved oxygen will be used up, causing the water to become deficient or even completely lacking in oxygen (anaerobic). This gives rise to fermentation and rotting processes, which cause a rapid deterioration in the quality of the water. The only way to ensure that the water retains enough oxygen is to limit the discharge of biodegradable organic matter.

P. F. Butzelaar and Drs L. P. J. Hoogeveen are with Philips Research Laboratories, Eindhoven.

Measurement of the oxygen demand

The most widely used method of measuring the amount of dissolved oxygen used up on the breakdown of organic matter is to determine what is known as the BOD_5^{20} value (Biochemical Oxygen Demand). This is done by measuring the amount of dissolved oxygen immediately after taking the sample, leaving the sample for five days at 20 °C under defined conditions and then determining the amount of dissolved oxygen again. The difference indicates how much dissolved oxygen has been taken up by the micro-organisms in the water during the biodegradation of organic matter (the BOD value). In good surface water the BOD value is 1-5 mg of oxygen per litre; in water from sewage-treatment plants it is often more than 20 mg/l. Although the information it supplies is very relevant, in that it reflects the processes taking place in the water itself, the method has many disadvantages, which include the following:

1) On average only 70% of the organic matter present will be broken down in the sample-testing period of five days [2].

2) The analysis also gives an incorrect indication if the water contains too little oxygen to break down all the organic substances, which may be the case when the concentration of the biodegradable matter is very high.

[1] A list of the parameters to be measured is given in the article by D. J. Kroon and M. Q. Mengarelli in this issue, p. 113.

[2] M. E. van der Harst, Proc. Machevo Congress, Utrecht 1965, p. 131.

The sample must then be diluted with oxygen-rich water, which means that the oxygen demand must be known approximately beforehand.

3) If the water contains toxic substances, the analysis may fail owing to poisoning of the micro-organisms. The same poisoning can of course occur in nature.

4) If the appropriate micro-organisms are absent or deficient, they have to be introduced.

5) Having to determine the dissolved oxygen twice takes up a great deal of valuable laboratory time and skilled labour.

6) Owing to the long time lag of the analysis the BOD_5^{20} method is not suitable for the control of treatment processes applied to effluents.

7) The result of the analysis will be reported too late in the event of accidents.

To offset these disadvantages to some extent the BOD methods are nowadays often supplemented by COD methods (Chemical Oxygen Demand) [3]. In these methods the biodegradable organic material is oxidized with a chemical agent instead of the dissolved oxygen, e.g. with potassium permanganate ($KMnO_4$) or potassium dichromate ($K_2Cr_2O_7$). This oxidation is of course done in a standard way, for example by boiling the water for two hours when $K_2Cr_2O_7$ is used [4]. The amount of chemical oxidant used yields a COD value. The COD values in good surface water are between 20 and 50 mg/l, and in water from sewage-treatment plants they may often be higher than 100 mg/l.

The COD methods previously employed do not offset all the disadvantages of the BOD methods. Whereas, for example, the standard boiling time is too short to oxidize all the substances, it nevertheless makes the total analysis time so long that the methods are not so suitable for process control in water-treatment plants. The detection limit of these methods is also fairly high. Chloride ions interfere with the measurement when $K_2Cr_2O_7$ is used, and a chloride determination must be made or $HgSO_4$ added to precipitate the chloride, or both. Finally, skilled personnel are required for the COD methods.

At Philips Research Laboratories a method has been devised in which a water sample is oxidized in a stream of nitrogen and oxygen (the 'carrier gas') at high temperature and in the presence of a catalyst, and the amount of oxygen used is measured with a zirconium-oxide cell [5]. The method overcomes the disadvantages mentioned above with regard to accuracy at low COD values: the speed of the measurement and the possibility of interference by chloride ions. The method is very suitable as a basis for automatically operating detectors designed for use both in automatic monitoring networks and in the control systems of treatment plants [*].

COD determination with a ZrO_2 cell

The ZrO_2 cell

The method we have developed for determining the COD value uses the stabilized-zirconia cell previously described in this journal [6]. The operation of the cell is based on the conduction produced in the ZrO_2 by negatively charged oxygen vacancies. When a carrier gas with a partial oxygen pressure p_i flows through the cell (see *fig. 1*), the partial oxygen pressure outside the cell being p_o and the cell with its electrodes being kept

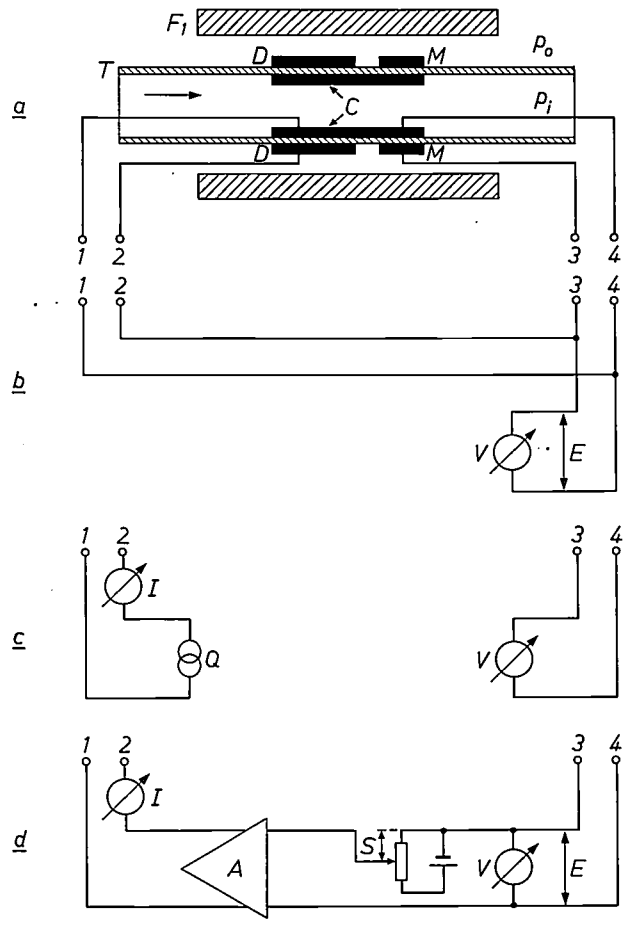


Fig. 1. *a*) Diagram of a zirconia cell for the measurement and control of oxygen pressures. *T* tube of zirconia stabilized with calcium oxide, through the wall of which negatively charged oxygen ions can be transported. F_1 furnace. *M* measuring electrode. *D* dosing electrode. *C* common inner electrode. The carrier gas is N_2. The ancillary equipment for various applications of the instrument is connected to terminals *1-4*.
b) Measuring circuit; a difference in partial oxygen pressure inside and outside, p_i and p_o, causes a potential difference *E* (measured with a voltmeter *V*) between the measuring electrode *M* and the common inner electrode *C*.
c) Dosing circuit; by means of a current source *Q* connected to the dosing electrode *D*, the inner electrode *C* and the ammeter *I*, oxygen can be fed into or extracted from the system, depending on the direction of the current.
d) Circuit for automatically controlling the partial oxygen pressure; by means of the variable voltage *S* and the amplifier *A*, a current can be produced that adds to or removes from the carrier gas an amount of oxygen such that the measured potential difference *E* maintains the same value as the set voltage.

at a temperature T, there will be a potential difference E between the inner electrode C and the measuring electrode M given by

$$E = \frac{RT}{4F} \ln \frac{p_i}{p_o}. \quad (1)$$

Here R is the gas constant, F the Faraday constant and 4 the number of electrons involved in the reaction $4e + O_2 \rightleftarrows 2\,O^{2-}$, causing the vacancies. If p_o is constant, e.g. equal to the partial oxygen pressure of the atmosphere, then E is a direct measure of the oxygen pressure p_i inside the cell.

If a current source is now connected to the inner electrode C and the dosing electrode D, oxygen will be fed into the system or extracted from it, depending on the direction of the current. The quantity of transported oxygen is given by

$$N = \frac{Q}{4F} = \frac{1}{4F} \int_0^t i\,dt, \quad (2)$$

where N is the number of gram molecules of oxygen, Q the quantity of charge delivered by the current source, and i the current. The effect of the dosing current can be determined by connecting a voltmeter between the measuring electrode and the inner electrode.

Since the quantity of oxygen supplied or extracted can thus be measured at the same time, a gas flow can be provided with a preselected oxygen content. The potential difference E for any given oxygen content can be calculated from equation (1). When the variable voltage S (fig. 1) is now set to the value corresponding to the required oxygen pressure, the amplifier A will deliver a current that feeds in or extracts a quantity of oxygen such that the measured potential difference E reaches the same value as the set voltage S. External changes of the oxygen pressure in the carrier-gas flow will cause corresponding changes in the current i, as required to keep the oxygen pressure at the set value. In the design of the amplifier special attention was of course given to the time constant of the ZrO_2 cell (0.1 to several seconds) and the logarithmic response, given by equation (1), to changes in the oxygen pressure.

The COD meter

Fig. 2 shows the schematic arrangement of the COD meter. The basis of the COD meter is a combustion furnace F_2. The carrier gas is conducted through the furnace in a quartz-glass column that contains the catalyst in the form of two platinum-gauze plugs. In the presence of the catalyst the pollutants in the water sample are burnt (oxidized) at a temperature of 900 °C with oxygen from the carrier-gas flow. The water sample (volume 10 μl) is injected into the glass column through a rubber membrane in the injection port *Inj*. The carrier-gas flow, kept constant by a flow regulator, is supplied with sufficient oxygen from a zirconia cell (the dosing cell, top left). The quantity of oxygen is adjusted with a variable current source Q_v.

After passing through the combustion furnace the gas enters a drying column D (usually containing $CaSO_4$) and passes through a second zirconia cell which, with its associated control circuit, acts as a measuring cell. The reference voltage in this control circuit is set in such a way, depending on the dosing current in the ZrO_2 dosing cell, as to cause a small zero current to flow to the dosing electrode in the measuring cell. When a water sample is now injected, the oxygen pressure in the carrier gas will drop as a result of the combustion of the injected organic matter. This changes the voltage between the measuring electrode and the inner electrode in the ZrO_2 cell, and the resultant difference between the measuring voltage and the set-reference voltage is amplified in the amplifier to produce a current i that adds exactly the amount of oxygen required to restore the original oxygen pressure. This amount of oxygen is exactly the amount used in the combustion of the injected organic matter. Integration of the current i with respect to time yields the number of coulombs from which the COD value can be calculated [7].

We shall now deal with a few points in somewhat more detail.

Complete combustion of the injected organic matter depends on sufficient oxygen being present at the

---

[3] In view of the relevance of the BOD analysis, the BOD value should continue to be determined in addition to a COD value. Moreover the BOD/COD ratio gives further information about the extent to which readily biodegradable matter exists side by side with matter which is difficult to break down biochemically. If this ratio is small it points to discharges of artificial substances that are not easily broken down biochemically, possibly from industrial effluents.

[4] Standard methods for the examination of water and waste water (U.S.A.), 1972.

[5] Particulars of the many other methods developed for determining the oxygen demand of water, and which are based on electrochemical principles, ultraviolet and laser Raman spectroscopy, gas chromatography and pyrolysis, will be found for example in:
L. Formaro and S. Trasatti, Anal. Chem. **40**, 1060, 1968.
A. P. Meijers, thesis, Delft 1970.
M. Mrkva, J. Water Poll. Control Fed. **41**, 1923, 1969.
N. Ogura and T. Hanya, J. Water Poll. Control Fed. **40**, 464, 1968.
E. B. Bradley and C. A. Frenzel, Water Res. **4**, 125, 1970.
T. S. Hermann and A. A. Post, Anal. Chem. **40**, 1573, 1968.
K. H. Nelson and I. Lysyj, Water Res. **3**, 357, 1969.

[6] N. M. Beekmans and L. Heyne, Philips tech. Rev. **31**, 112, 1970.

[7] When stating the COD value the method used to determine it should also be mentioned. This can be indicated, for example, with a subscript. Thus, our method is generally referred to as the COD_{ZrO_2} method, or even as the TOD_{ZrO_2} method (Total Oxygen Demand), in view of the virtually complete oxidation obtained with it.

[*] Detectors of this type are being developed by the Philips Scientific and Analytical Instruments Group.

moment of injection. It is of course possible to use a carrier gas with a high oxygen pressure, so as to ensure that enough oxygen will be present in all conceivable circumstances. Should the sample contain little combustible matter, the relative decrease in the oxygen content will be small and, owing to the logarithmic response of the ZrO_2 cell, the measured change in potential difference will be extremely small, reducing the accuracy of the method. If, however, the oxygen pressure in the carrier gas is variable, it is always possible to keep the oxygen pressures in such a ratio that a sufficiently large potential difference will occur. In the COD meter discussed here this adjustment of the oxygen pressure is made by means of the dosing cell, which is easily regulated. This implies incidentally that unknown samples may often have to be measured a second time, with a first measurement to estimate the optimum oxygen pressure, and a second to give the actual accurate determination.

If all organic constituents of the water are to undergo complete combustion, they must remain sufficiently long in contact with the oxygen and the catalyst in the combustion furnace. On the other hand, it is desirable to keep the analysis time as short as possible. A carrier-gas flow rate of about 60 ml/min ensures a sufficiently long contact duration for an analysis time of only 1.5 to 2 minutes.

Results from the literature [8] and those of our own show that very fast and complete combustion of nearly all organic substances takes place at a temperature above 850 °C. As will appear below, the only problems are found with nitrogen compounds.

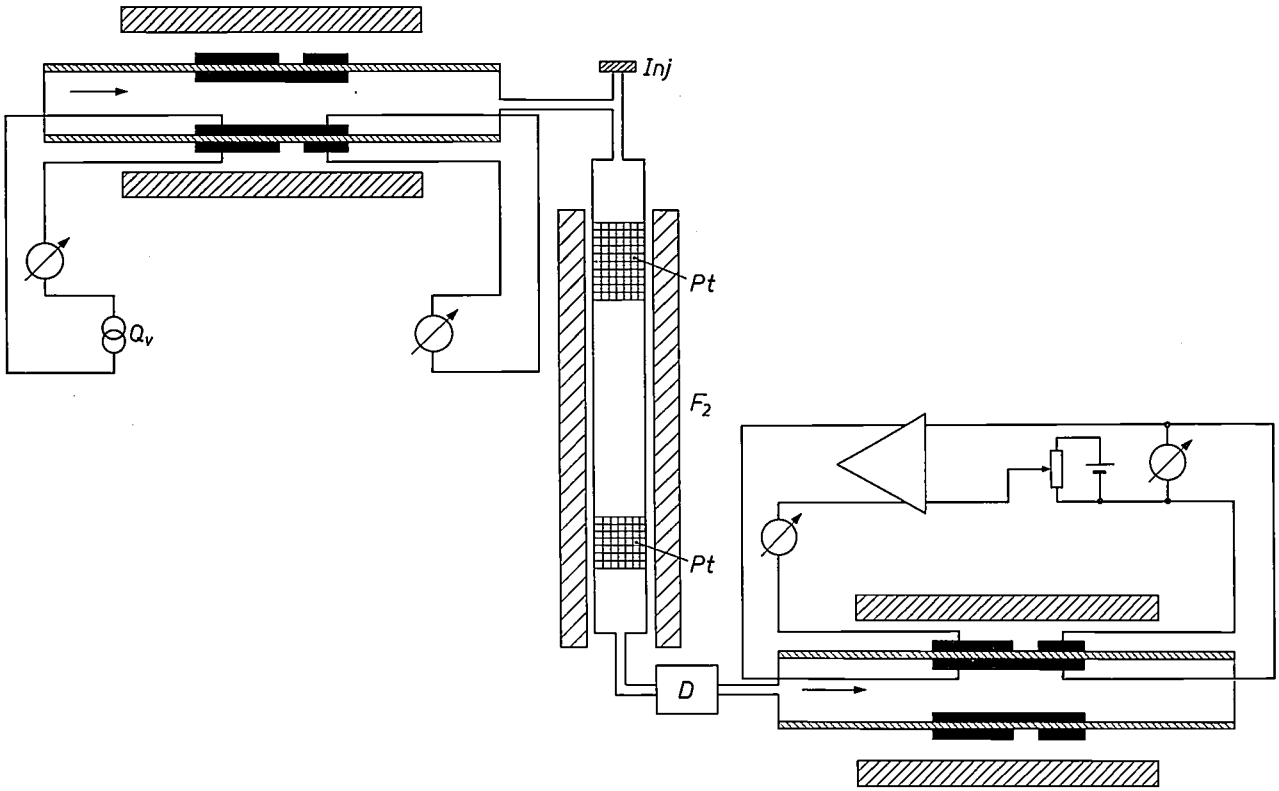


Fig. 2. Diagram of a COD meter using a zirconia cell. *Left*: a ZrO_2 cell with a variable current source Q_v which delivers a constant O_2 pressure to the carrier gas. *Right*: a similar cell that acts as a measuring instrument. The heart of the instrument is a combustion furnace F_2, which contains two plugs of platinum *Pt*. The water sample (10 µl) is injected through *Inj*. *D* is a drying column filled with $CaSO_4$. The variable voltage in the second cell is set in such a way that this cell tends to maintain the oxygen content delivered by the first cell. The consumption of oxygen in the combustion of the water sample is compensated by the dosing current in the second cell.

Features of the COD meter

Linearity

The linearity of the COD meter was checked with standard solutions of sodium acetate in distilled water, which had COD values of 50, 100, 150 and 200 mg/l. As can be seen from *fig. 3*, the signal obtained in the region of interest is proportional to the sodium-acetate concentration. It also appears that the signal obtained with the standard solutions is virtually identical with the calculated signal. In practice the proportionality is valid up to COD values of about 3000 mg/l, provided

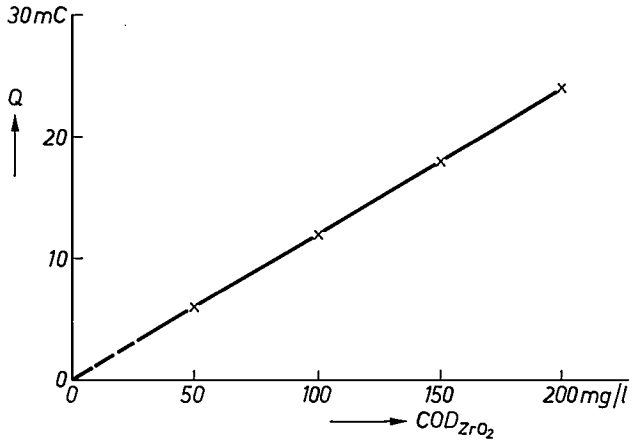


Fig. 3. The charge Q measured by a zirconia cell as a function of the calculated COD values of standard solutions of sodium acetate in distilled water. As can be seen, the measured signal is proportional to the COD value in the concentration region of interest.

Table I. Combustion efficiencies based on the chemical equation $C_nH_mO_pN_x + (n + m/4 - p/2)O_2 \rightarrow nCO_2 + m/2\ H_2O + x/2\ N_2$ of various organic compounds at concentrations of 50 mg/l and 150 mg/l. The excessive efficiency in the case of glycine points to partial conversion of the active amino group into NO.

	COD 50 mg/l	COD 150 mg/l
Benzoic acid	—	100%
Phenol	101%	100%
Citrate	—	104%
Oxalic acid	106%	104%
Glycine	126%	118%
Nitroaniline	108%	105%
Ammonium chloride	106%	—
Urea	100%	—
Aniline	—	108%

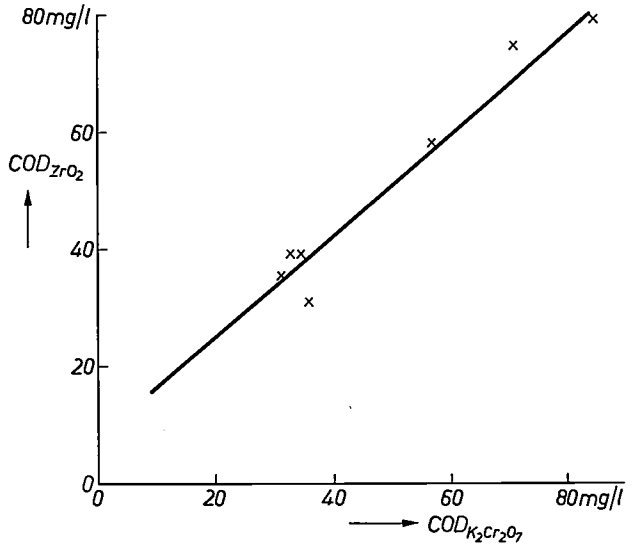


Fig. 4. COD values obtained by the method described here, compared with the method using $K_2Cr_2O_7$ as oxidant. The water samples were taken from small rivers in the province of North Brabant, the Netherlands. From the fact that the curve does not pass through zero it may be inferred that the ZrO_2 method involves more complete oxidation than the $K_2Cr_2O_7$ method.

the gain of the control circuit is sufficiently large and the oxygen pressure in the carrier-gas flow is sufficiently high.

Apart from the standard solutions of sodium acetate in water, standard solutions were measured of a number of organic substances with oxygen atoms in the organic molecule that would be expected to affect the behaviour during combustion. *Table I* shows that there is good agreement between the measured and the theoretical values for the compounds containing only carbon, hydrogen and oxygen (benzoic acid, phenol, sodium citrate and oxalic acid).

Compounds containing nitrogen in the form of an NH_2 group such as glycine use up more oxygen than is calculated from the reaction equation

$$C_nH_mO_pN_x + (n + m/4 - p/2)O_2 \rightarrow nCO_2 + m/2\ H_2O + x/2\ N_2.$$

This may be explained by assuming that the amino group is partly converted into NO.

Similar considerations apply to such substances as ammonium compounds. Further investigations will have to reveal how this partial conversion into NO depends on the oxygen pressure, temperature, time spent in the column (involving further oxidation to NO_2), on whether or not the nitrogen group is activated in the original organic compound, and so on.

Interference

In principle all constituents of a sample that give off oxygen in the defined conditions can make a negative contribution to the COD value. In practice nitrates are the principal source of interference, being converted in the combustion furnace into oxygen and nitric oxide. The oxygen thus released will of course be used up in the combustion of other substances, so that in fact the COD value found is too low. Analysis of the efficiency of the reactions that take place shows that nitrate in concentrations of up to about 400 mg/l yields 1 mol of oxygen per mol of nitrate. Since the concentrations of nitrates found in surface water are fairly low (a few mg/l to about 20 mg/l at the most) this interference is unimportant. (Even at a concentration of 20 mg/l the interference on the COD value is only —10 mg/l.)

Accuracy; comparison with the $K_2Cr_2O_7$ method

Fig. 4 shows the results of measurements on a number of samples taken from various small rivers around Eindhoven. The COD values determined from oxidation with potassium dichromate as described above ($COD_{K_2Cr_2O_7}$) are compared with the COD values de-

[8] V. A. Stenger and C. E. Van Hall, Anal. Chem. 39, 206, 1967.

termined by the method described in this article (COD_{ZrO_2}). The analyses by the $K_2Cr_2O_7$ method were done in the laboratory of the Water Board for the river Dommel. These results and analyses of a large number of samples taken from other rivers (the Rhine and the Maas) show that

$$COD_{K_2Cr_2O_7} \approx 0.9\ COD_{ZrO_2}.$$

This may be attributed to the more complete oxidation that takes place at high temperature in the COD_{ZrO_2} method.

The experiments with standard solutions show that an uncertainty of 2 to 3% can be achieved with our method of COD determination. This is not less than the uncertainty obtained with the standard analysis method based on oxidation with $K_2Cr_2O_7$, where our method has a distinct advantage in the very much shorter time required for analysis (2 minutes as against 2½ hours).

In samples taken from surface water and from water that has passed through a treatment plant, the spread in the results of the measurements both by the $K_2Cr_2O_7$ method and by our method increases to about 10% because of the presence of sedimentation sludge, suspended matter, etc. with the substances adsorbed to it. This is a clear illustration of the problem of representative sampling.

Because of the short analysis time and the ease with which it can be automated, the method we propose can be applied in an automatic monitoring network both for measuring trends and for giving a warning signal if the norm is exceeded.

Summary. The COD value is determined by measuring the quantity of oxygen used up when a sample of water is oxidized at 900 °C with oxygen in a constant carrier-gas flow in the presence of platinum. The oxygen is supplied and its partial pressure kept constant by a zirconia cell. The carrier gas is also passed through a second zirconia cell, which is set up by means of a control circuit in such a way that it brings the oxygen content to the value prevailing at the output of the first cell. The electric current that restores this value is a measure of the quantity of oxygen used up in the oxidation of the water sample. The total duration of the analysis is very much shorter than in other methods (about 2 minutes as against a few hours). The method is easily automated and lends itself to application both in monitoring networks and in control circuits for water-treatment plants.

Polychrome data display using a single TITUS tube

The optical relay tube of the TITUS type (*Tube Image à Transparence Variable Spatio-temporelle*) was primarily developed for the large-screen projection of television pictures [1] [2]. At the present time a luminance of 32 cd/m² can be obtained on a screen area of 25 m²; this corresponds to a total luminous flux of 2500 lm. We shall explain below how this tube can be used for the display of data in different colours. Before doing so it will be useful to recapitulate the operation of the TITUS tube.

As illustrated in *fig. 1*, the TITUS tube is used to produce a modulation in time and space, corresponding to the picture to be projected, on the beam of light from a powerful external light source L. This is done by employing the Pockels effect — a longitudinal electro-optical effect — of a crystal plate C of KD_2PO_4 (potassium dideutero-orthophosphate), a ferroelectric material that crystallizes in a uniaxial lattice. The manner in which the image information is introduced into the crystal plate can be described by imagining the crystal to be divided into a large number of elementary capacitors that have one electrode in common, which is a conducting layer E on the front of the plate. The video signal is applied to this electrode. The back of the plate is scanned by an electron beam of constant intensity, giving rise to secondary emission. Since a grid G at a fixed potential is situated at a short distance from C, the electron beam acts as a flying short-circuit between C and G, so that the scanned capacitive element of the crystal plate is charged or discharged depending on the video voltage at the instant of scanning.

In the charged state the element in question gives the Pockels effect — an induced birefrigence proportional to the electric polarization εE. The crystal has its optical axis (the z-axis) perpendicular to the plane of the plate. The direction of the electric field coincides with this. Thus, if a light beam polarized in one of the other crystal directions by a polarizer P_1 (e.g. in the x-direction, in the plane of the crystal) passes through the crystal element, it will be split into two components polarized in the directions of the bisectors of the angles between the x- and y-axes and propagating at different velocities.

After the light beam has passed through the crystal element — *twice* in the TITUS tube, since it is reflected by a dielectric mirror M situated behind C — a phase difference ϕ appears between the components. If the beam now passes through a second polarizer P_2, which is crossed in relation to P_1, the resultant transmission

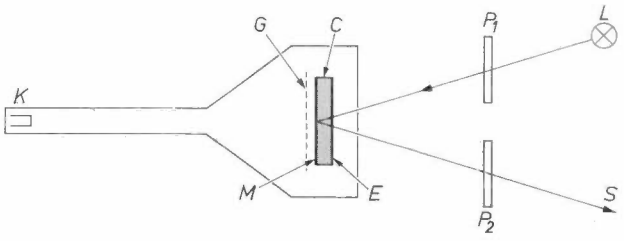


Fig. 1. Basic structure of the TITUS tube. The cathode K at a potential between −500 and −1000 V provides an electron beam of constant intensity. C monocrystalline plate of KD_2PO_4 giving the Pockels effect (birefringence dependent on the longitudinal electric field-strength). M multilayer dielectric mirror. G earthed collector grid placed a short distance away from the crystal plate. E transparent conducting electrode receiving the video signal. For practical purposes the electron beam can be regarded as a flying short-circuit between the grid and the target. L light source. P_1 and P_2 crossed polarizers. The light is projected on to the screen S.

will be proportional to $\sin^2 \frac{1}{2}\phi$ [3]. The phase difference ϕ is proportional to the product of the field-strength in the crystal element and the thickness of the plate. In this way the light beam is modulated in intensity at the position of the element by the video signal at the instant of scanning.

An important characteristic of the TITUS tube is that the crystal plate is cooled to a temperature near the Curie point of the material. This temperature depends on the deuterium content and is −55 °C if deuterium is substituted for 95% of the hydrogen in KH_2PO_4. The low temperature has two advantages. Firstly, in this temperature range the Pockels effect is considerably greater and a very much lower modulation voltage is required than at other temperatures. Secondly, the projected images are free from flicker, since the crystal elements cannot discharge at this temperature in the time between two scans. In fact the time constant is so large that the tube can be used as a storage tube for periods of about a quarter of an hour, e.g. in computer peripherals.

To project colour television pictures three TITUS tubes are needed, but if no more than a few colours are needed, e.g. for a data display, it can be done with a single tube. Our method takes advantage of the fact that the ultimate amplitude modulation of the light beam is obtained from phase modulation in the crystal.

[1] G. Marie, Un nouveau dispositif de restitution d'images utilisant un effet électro-optique: le tube TITUS, Philips Res. Repts. 22, 110-132, 1967.
[2] G. Marie, Large-screen projection of television pictures with an optical-relay tube based on the Pockels effect, Philips tech. Rev. 30, 292-298, 1969.
[3] See for example E. E. Wahlstrom, Optical crystallography, 4th edition, Wiley, New York 1969, from page 157.

The phase difference of the two components into which the incident light beam is split is given by

$$\phi = \frac{2\pi l \Delta n}{\lambda}. \qquad (1)$$

The quantity l is equal to twice the thickness of the plate, λ is the wavelength of the light and Δn the difference, dependent on the video voltage, between the refractive indices in the two new directions of polarization. When crossed polarizers are used the transmission T is given by

$$T = \sin^2 \tfrac{1}{2}\phi = \sin^2 \frac{\pi l \Delta n}{\lambda}. \qquad (2)$$

Fig. 2 shows the transmission as a function of the path difference $l\Delta n$ for three light components (red, yellow and green). It is clear that in the region $l\Delta n < 200$ nm the three colours are transmitted more or less in proportion. If the incident light is white, the emergent modulated beam will also be practically white, with slightly too little red and too much blue. This region *B-W* is thus suitable for monochrome projection. However, if we wish to convert the amplitude modulation of the light into a colour modulation, we must choose an operating region in which variation of the video signal will give large differences in the intensity of the desired colour components in the transmitted light. We have therefore used a phase plate, in series with the tube, which adds a fixed path difference between the emergent beam components to the path difference already present in the tube due to the Pockels effect. In the example given in fig. 2 a polychrome operating region (*Y-R-G*) is thus created by increasing the path differences by about 950 nm. When a filter is used to remove the blue light from the spectrum of the incident beam, the hue of the transmitted light varies from yellow through red to green for signals ranging from 0 to about 180 V. If a fixed signal level is used for displaying particular items of data, these will then invariably be projected in the same colour.

The variation in colour will be explained in more detail with the aid of *figs. 3* and *4*. Fig. 3 shows the way in which the spectral composition of the projected light varies in the colour triangle when the path difference $l\Delta n$ is increased from 915 nm (point *1*) to 1240 nm (point *14*). Curve *I* refers to white incident light and curve *II* to the same light, but with the blue filtered out. Fig. 4 gives a clearer impression of the projected colours and their relative intensity L obtained when the path difference is varied. When a phase plate giving a path difference of 1115 nm is used, the range of signal-voltage variation is from −128 V to +80 V. This constant path difference corresponds to minimum luminance at zero voltage. The colours with these low rel-

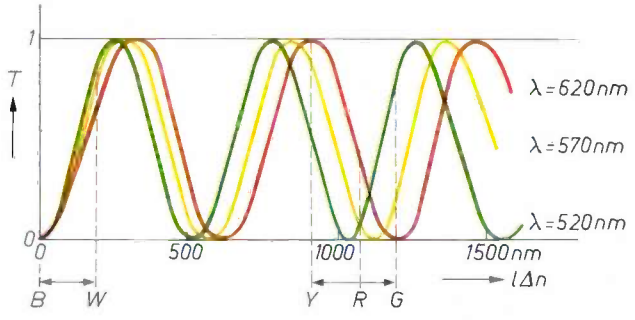


Fig. 2. Transmission T of a birefringent crystal between crossed polarizers as a function of the path difference $l\Delta n$ between the two differently refracted light components, for red, yellow and green light. The region *B-W* is suitable for monochrome projection, the region *Y-R-G* is a polychrome operating region.

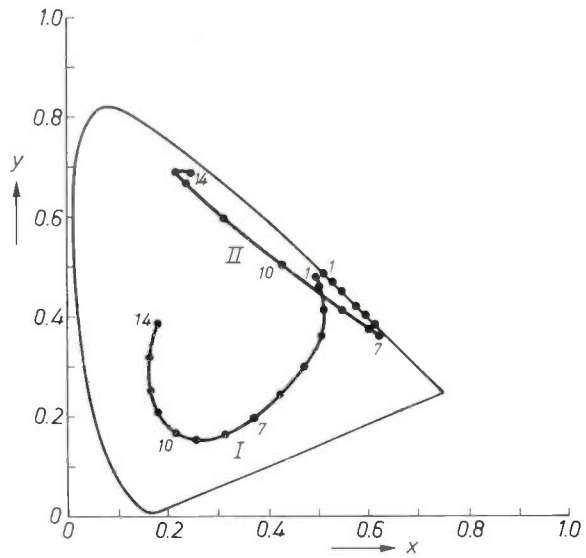


Fig. 3. Colour triangle showing the locus of the colour coordinates of the transmitted light when the path difference $l\Delta n$ is varied from 915 nm (point *1*) to 1240 nm (point *14*). The path difference between neighbouring points is always 25 nm. Curve *I* relates to white incident light, curve *II* to a spectrum with wavelengths greater than 500 nm (blue light filtered out).

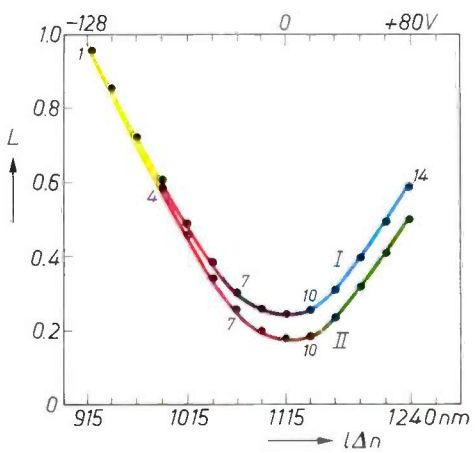


Fig. 4. Relative intensity L and colour of the transmitted light as a function of the path difference $l\Delta n$ and the signal voltage V on the TITUS tube, with a path difference of 1115 nm introduced by the phase plate. Curve *I*: white incident light. Curve *II*: incident light containing only wavelengths above 500 nm. The curves are derived from the colour coordinates shown in fig. 3.

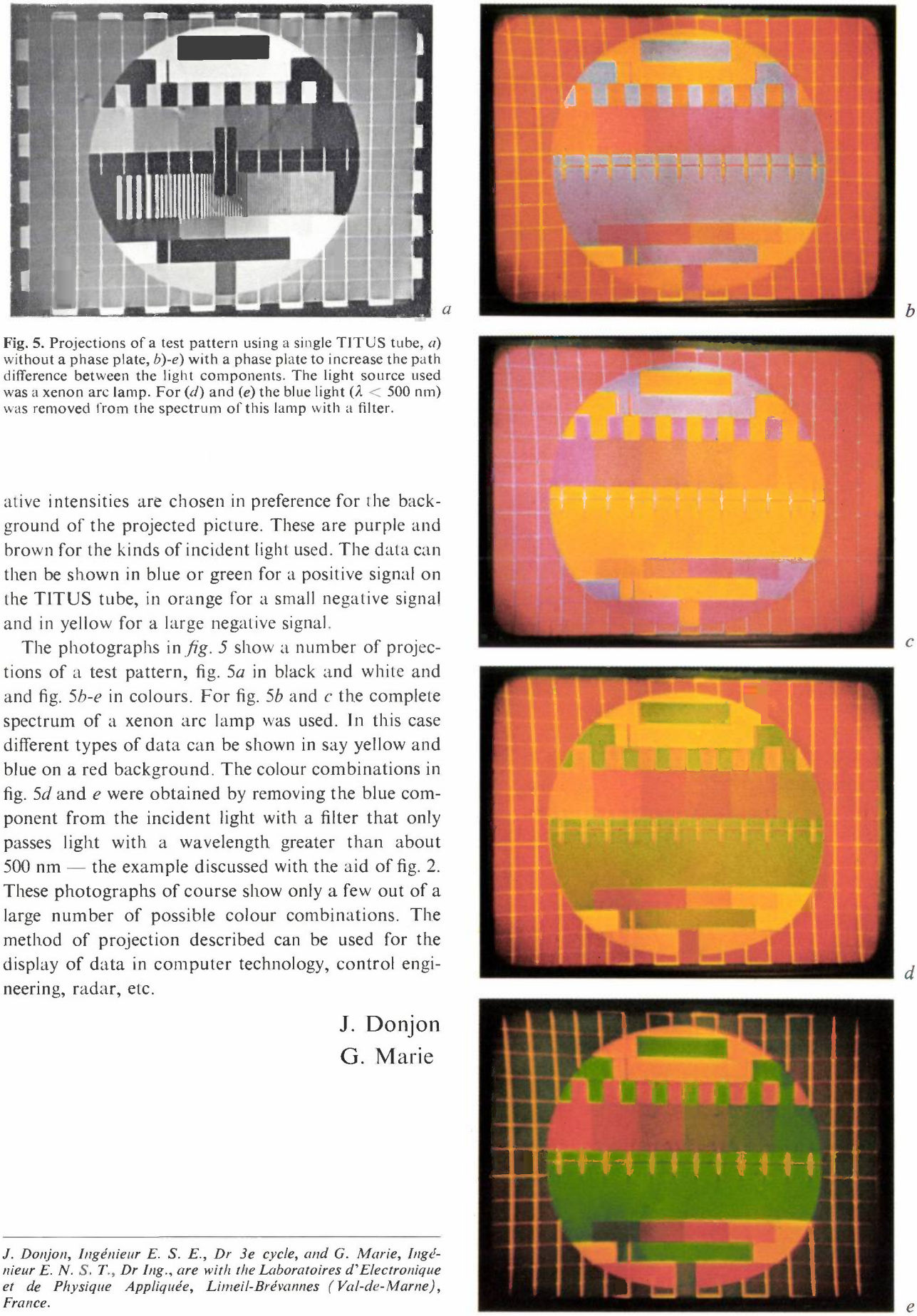


Fig. 5. Projections of a test pattern using a single TITUS tube, *a*) without a phase plate, *b*)-*e*) with a phase plate to increase the path difference between the light components. The light source used was a xenon arc lamp. For (*d*) and (*e*) the blue light ($\lambda < 500$ nm) was removed from the spectrum of this lamp with a filter.

ative intensities are chosen in preference for the background of the projected picture. These are purple and brown for the kinds of incident light used. The data can then be shown in blue or green for a positive signal on the TITUS tube, in orange for a small negative signal and in yellow for a large negative signal.

The photographs in *fig. 5* show a number of projections of a test pattern, fig. 5*a* in black and white and and fig. 5*b*-*e* in colours. For fig. 5*b* and *c* the complete spectrum of a xenon arc lamp was used. In this case different types of data can be shown in say yellow and blue on a red background. The colour combinations in fig. 5*d* and *e* were obtained by removing the blue component from the incident light with a filter that only passes light with a wavelength greater than about 500 nm — the example discussed with the aid of fig. 2. These photographs of course show only a few out of a large number of possible colour combinations. The method of projection described can be used for the display of data in computer technology, control engineering, radar, etc.

J. Donjon
G. Marie

J. Donjon, Ingénieur E. S. E., Dr 3e cycle, and G. Marie, Ingénieur E. N. S. T., Dr Ing., are with the Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes (Val-de-Marne), France.

Philips tech. Rev. 34, 132-145, 1974, No. 5/6

The microprogram control of the Philips P1000 family of computers

J. A. Dinklo and E. B. de Vries

Articles describing the operation of computers do not usually go into much detail about the way in which their operation is controlled. We refer here to the mechanism that causes the various elementary operations forming a machine instruction to be performed in the correct sequence. A control counter can be used for this, but a more flexible method has been developed, called microprogram control. This article describes the microprogram control used in the computers of the P1000 family, which differs from that used in other computers, particularly in the method of making jumps in the microprogram.

Introduction

In general terms a computer can be considered as a collection of input and output units, memories, a central processor and a control unit [1]. The *input and output units* serve for the input and output of the information. Sometimes they operate at a distance by way of a data-communication link from 'terminals': teletypewriters, or image displays with a keyboard. The *memories* (sometimes called stores) store numbers and programs. Memories of several different kinds can often be found in the same computer: internal and external stores, working memories and backing stores each with its own characteristics for speed, capacity, physical principle employed, methods of use and last but not least, its cost. The *central processor* ensures that the desired operations are performed upon the information. In its simplest form the central processor consists of two registers and an arithmetic unit that can perform an operation (e.g. an addition) on the contents of the registers, and then transfer the result back into one of these registers or into another register in the computer. The *control unit* ensures that the computer executes the programs stored in the memory without further attention. The control unit is often considered to be part of the central processor, and we shall follow this convention here.

In this article we shall be considering the operation of control units based on a '*microprogram*'. This method of control has a number of advantages over the *counter control* originally used; indeed, if a family of computers is to be designed, microprogramming is in effect the only possibility. Our treatment of this subject will be based on the control used in the computers of the Philips P1000 family, with special attention to the control in the P1400. We begin by recapitulating a few general concepts that are of significance in computer control.

Ir J. A. Dinklo and E. B. de Vries are with the Philips Data Systems Division, Apeldoorn, the Netherlands.

Elementary operations in a computer

The computer programs are present in the memory in the form of a series of 'machine instructions', simple operational steps such as:

— 'fetch a number from the memory and store it in register A of the central processor', or:
— 'fetch a number from the memory and add it to the contents of register A', or:
— 'jump to another point of the program'.

Each machine instruction of this type usually requires the execution of a number of elementary operations, either simultaneously or in sequence. For the instruction of the first example these could be reading the number out of the memory and — in the case of a core store — writing it back into the memory again, transferring it to register A and selecting the following instruction and fetching it from the memory.

Programmers do not usually work in 'machine instructions' nowadays, but in a 'high-level language'. This high-level language has to be converted into machine instructions before the machine can start to execute the program. This conversion can be made by the machine itself with a special program known as a 'compiler'.

[1] Earlier articles in this journal giving a detailed description of a computer are:
W. Nijenhuis, The 'PASCAL', a fast digital electronic computer for the Philips Computing Centre, Philips tech. Rev. 23, 1-18, 1961/62;
G. J. A. Arink, The onboard computer of the Netherlands astronomical satellite (ANS), Philips tech. Rev. 34, 1-18, 1974.

In all these operations information has to be transferred from one register to another register; so that this transfer of register contents is one of the basic operations in a computer. The transfers are made by gates, see *fig. 1*. If a condition signal *BYA* is present, the contents of register *A* are copied into register *B* at the instant a clock pulse *Cl* appears. The contents of the register *A* remain unaffected by this. The condition signal *BYA* is in principle only present during a single clock pulse. In simple terms all that the control unit has to do is to arrange that the condition signals that permit the desired transfers of information are generated at the correct instants in time. The generation of such a condition signal for a particular transfer can be considered as the most elementary operation in the machine. The generation of read or write signals for the working memory can also be considered as such elementary operations.

Modification and indirect addressing

Even the execution of an apparently simple instruction such as 'fetch the number at address *n* from the memory and put it in register *A*' is performed in a much more complicated way in a modern computer than might at first be thought. This is because the address part of an instruction does not necessarily refer directly to the actual address in the memory. First of all, the programmer can *modify* the address. This is a method of addressing in which the actual address is obtained by adding the contents of an *index* register to the given address. Address modification is a useful aid in operations such as the addressing of a series of numbers in a row of successive memory locations, e.g. the components of a vector $a(i)$ ($i = 0, 1, 2, \ldots$). The programmer can then use one and the same instruction to reach all the components of the vector by putting the number of the desired component into the index register; if the complete series has to be processed he then has only to increase the contents of this register by 1 each time. The execution of such a modified instruction does of course require one extra addition.

The programmer can also quote an *indirect* address,

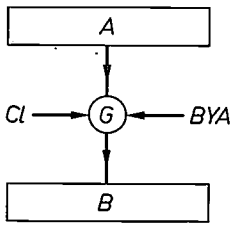


Fig. 1. Principle of the most widely encountered elementary operation in a computer. The contents of register *A* are transferred into register *B* at the first clock pulse *Cl* to appear after the condition signal *BYA* has been applied to the gate *G*. The condition signal is produced by the control unit.

the given address does not then contain the desired number, but the location at which it is to be found (i.e. its address). Such an instruction requires two readings from the memory instead of one.

The address produced by modification or indirect addressing of the address part will from now on be called the *relative address*.

Multiprogramming

In practice the situation is even more complicated since a large computer is usually handling a number of different and unrelated programs at the same time (multiprogramming). If for one reason or another no progress can be made with a program that is being run, the machine switches for a while to another. If something appears, say from a terminal, that has priority, the program being run must be interrupted; this is an *interrupt*.

Since the space available in the working memory is limited, it is sometimes necessary to store the first program temporarily in a backing store, and to bring the other one out again. Very long programs sometimes have to be run in *segments* — separate sections of the program that are more or less independent and can be transferred individually from the backing store to the working store as required — so that the same situation also holds for these segments of programs. In this way, every appearance of a program may be at different locations in the working memory; the 'relative addresses' therefore have nothing do to with the actual or absolute addresses in the memory during the execution of the program.

The execution of programs in a modern computer is therefore always monitored by a *housekeeping routine*. This notes exactly which program is being run, where it is in the working memory, which programs are waiting and where they are stored. Whenever a program or segment is transferred to the working memory for execution, the housekeeping program ensures that it is put into an empty part of the memory. The initial or 'base' address is noted by the housekeeping program. While the program is being run each absolute address is then obtained by adding this base address to the relative address. At the same time a check can be made to ensure that no error in programming has put an address into an area already occupied by another program. If this were to happen the other program could unwittingly be written over and erased or perhaps merely read.

Because of these various complications in addressing the execution of an instruction takes place in more stages than was originally indicated above. The control unit therefore has to put a fairly large number of elementary operations under way for each instruction.

Microprogram control

In older machines, and also in modern machines when high speed is required, the initiation of the desired elementary operation is controlled by the *control counter*, a sequential circuit that can operate cyclically through a (usually large) number of steps. In principle such a counter (*fig. 2*) consists of a register R whose contents can be considered as the state of the counter. The register is connected to a logic circuit S_1 designed in such a way that applying the contents of the register as the input signal causes the following state to appear at the output. When a clock pulse *Cl* opens the gate, this new counter state is taken over by the register. In this way the counter makes a step in its cycle at each clock pulse. A second logic circuit S_2 derives condition signals *Co* from the state of the counter; the signals *Co* initiate the associated elementary operations. The complete counting cycle corresponds to a single machine instruction. Since different instructions often require different numbers of elementary operations, then either these different numbers of operations must be fitted to the same pattern of counting cycle, or the counter must be in some way variable or adjustable, or both. *Fig. 3* shows a diagram of the states of a very simple adjustable counter with eight states. In state *3* the counter awaits a signal *A* or *B*. If signal *A* appears, the counter goes from state *3* to state *0*; if *A* does not appear but *B* does, then the counter goes from state *3* to state *4*. In a similar way at state *7* the counter awaits a signal *C* to appear before going on to state *0*. The signals *A*, *B* and *C* are applied to the logic circuit S_1 of fig. 2.

In a counter of this type, which is usually considerably more complicated than the one we have described, the logic is usually fairly complicated. Since this logic is built into the machine in the form of permanently connected circuits ('wired logic'), each type of machine is physically unique, and once the design and construction is completed it is almost impossible to modify a machine or to extend its range of machine instructions.

As far back as 1951 M. V. Wilkes [2] showed how to organize this counter logic in a more systematic and generalized manner, which would at the same time permit a computer to be modified more easily. He described what he called a 'microprogram organization', in which the successive elementary operations for each instruction were laid down as a short program. Although at the time he was thinking of a version using diodes, he already saw the possibility of putting the microprogram into a store, as is usually the case today. The memory used for this is frequently of the 'read-only' type: information is written into it only once, during manufacture, and can only be read during operation, it cannot be altered. This kind of memory is often less expensive and is usually faster.

When a microprogram memory or *control store* is used the arrangement is as follows (*fig. 4*). The contents of the register R determine an address in the control store *CS* via an address-selection circuit S_1. The microprogram word located at this address is read and transferred to the buffer register U. Each word consists of two parts: an *address part*, also known as a sequence part (*seq*), and a *command part* (*com*). The address part determines the next contents of the register R, in simple cases by the transfer of this address part to the register, and sometimes in a more complicated way as we shall see later. In the command part each bit or group of bits (called a field) corresponds to a particular elementary operation or group of operations. The corresponding flipflops (bistables) in the buffer register U provide the condition signals for these operations. The command part of U thus carries out the function of the logic circuit S_2 in counter-organized control (fig. 2).

In this microprogram-organized control, the execution of a single machine instruction, which in counter-organized control corresponded to a single counting cycle, corresponds to the execution of the instructions in a particular row of microprogram words. The control unit now has to be built in such a way that it can

— read a microprogram word and put it into the buffer U,
— initiate the elementary operations given in the command part, and
— convert the sequence part into the address of the next word in the control store.

The operation is now uniform: the way in which a machine instruction is executed depends only on the contents of the control store. The complete organization is an example of 'stored logic' instead of the 'wired logic' of the previous solution [3].

In a computer with microprogram control it is relatively simple to add extra functions to the machine at a late stage of construction. It is also possible to test the operation of the machine *before* it has been built, with the aid of special programs. Microprogramming also contributes towards greater reliability, because wiring becomes simpler. Maintenance is also simpler when microprogramming is used; if a fault appears the blocks of the control store can easily be changed. In addition, it is about the only reliable method for producing a family of computers, and for emulating other machines.

[2] M. V. Wilkes, The best way to design an automatic calculating machine, Manchester University Computer Inaugural Conference 1951, page 16.

[3] See also for example:
S. S. Husson, Microprogramming — Principles and practices, Prentice-Hall, Englewood Cliffs, N.J., 1970;
A. J. van de Goor, Microprogrammeren, Informatie **15**, 380-384, 1973 (No. 7/8).

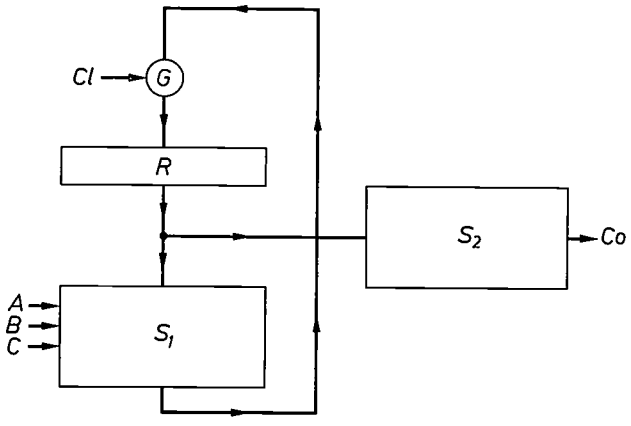


Fig. 2. Principle of a control counter. The contents of the register R are called the 'state' of the counter. S_1 is a logic circuit that for any counter state applied to its input gives the next state as its output signal. This state is taken over by the register via the gate G at the next clock pulse Cl to appear. Extra condition signals (A,B,C) affect the sequence of the various states (see fig. 3). S_2 is a logic circuit that derives the required condition signals Co from each counter state.

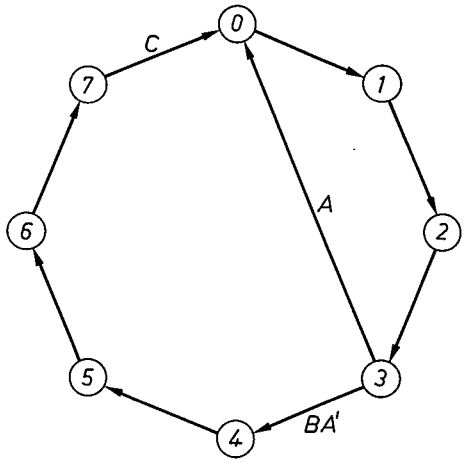


Fig. 3. Diagram showing the states of a simple control counter. The counter recognizes eight states. In state *3* the counter awaits the appearance of a condition signal A or B. If signal A appears, the counter jumps from state *3* to state *0*; if A does not appear but B does, then the counter goes to state *4*. In state *7* the counter awaits a signal C.

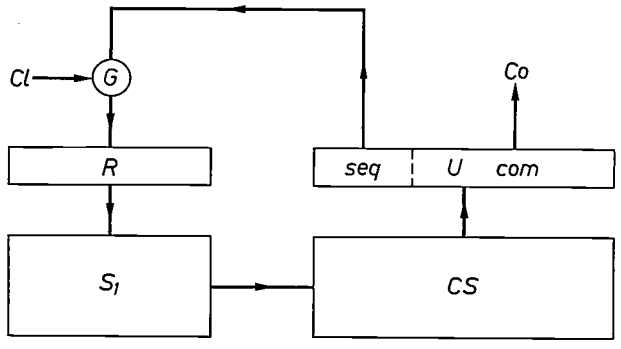


Fig. 4. Diagram to illustrate the principle of microprogram control. The contents of the register R determine an address in the control store. This address is selected by the selection circuit S_1, and its contents are then read and put into the buffer register U. The word is divided into an address or sequence part *seq* and a command part *com*. The address part determines the address of the next microprogram word, which is transferred to the register R at the next clock pulse. The command part delivers the condition signals Co.

Emulation, simulation and interpretation are terms often encountered today in the computer literature. We shall attempt to explain them here. The facility of being able to 'imitate' one machine (A) on another one (B) is often very useful. For example, an impression of the behaviour of A can be obtained and programs intended for A can be run unchanged on B.

This 'imitation' can in principle be done in three ways: simulation, interpretation or emulation. In simulation a program for B is written whose result gives information about specific points of interest in the behaviour of A, and without imitating the behaviour of A in every detail. In interpretation each machine instruction of A has a short piece of program for B that will produce the same effect, and a program for A is then run on the machine B by executing the short pieces of B-program associated with the A-instructions.

In emulation special pieces of microprogram are put into the microprogram memory of B to simplify the simulation. In the most detailed case this process can go so far that the complete microprogram package for the A-instructions is built into the machine B. It is a very efficient method that is technically relatively simple, particularly in the case of a read/write microprogram store.

Reading the microprogram words usually takes a longer time than a logic circuit takes to deliver an output signal. For the same speed the cost of microprogramming is therefore higher than that of counter control. Nevertheless, the continued decrease in the cost of computer stores has encouraged the use of microprogramming in almost all computers; the method is too slow only for the very fastest machines.

Since we want to discuss the microprogramming of the computers of the P1000 family we should first look at some of the characteristic features of these computers.

The P1000 family of computers

The P1000 family of computers includes the models P1075, P1100, P1175, P1200 and P1400. The lower numbers refer to the simpler 'smaller' machines and the higher numbers to the 'larger' ones. All these models are compatible with one another, which means that they have the same instruction repertoire and that the housekeeping routines and all the other programs written in machine language for a smaller machine can always be used on the larger models. The converse will of course only hold provided that limitations of the configuration are taken into account. Such computers are said to have the same hardware/software interface or the same system architecture.

The differences between the models are chiefly in performance, i.e. the data-processing rate, the capacity of the available internal memories, and the number of peripheral units that can be connected. Since the basic electronic 'building blocks' are the same for the various

models, this is achieved by increasing use of parallel operation in the faster and larger models.

The instructions must have exactly the same effect in the different models, even though they may be executed at different rates. The P1075 computer, for example, operates with a data path of 8 bits (an 8-bits-wide or eight-fold path), i.e. 8 bits are transferred simultaneously, while the P1400 has a data path of 32 bits (32-fold path). An operation in the arithmetic unit, such as the addition of two numbers of 32 bits will therefore always take four times as long in the P1075 as in the P1400, since this operation in the P1075 can only be done in steps of 8 bits.

Fig. 5 shows the principle of the models in the P1000 series. The most important units are the working memory *MEM* and the central processor *CPU*, with the control store *CS* considered here as a subunit of the central processor. The peripherals *P* are connected via a 'channel' *CH*. This term 'channel' refers to a unit that exercises a nearly autonomous control over the data flow to and from a peripheral unit. The peripherals each have their own control units, and sometimes their own control stores (again replacing a counter-control arrangement).

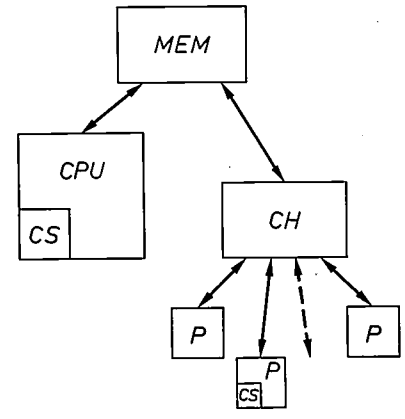


Fig. 5. Basic P1000 arrangement. *MEM* working memory. *CPU* central processor, with the control store *CS* shown as a subunit. *CH* 'channel', an autonomous unit controlling data flow to and from the peripherals *P*. Some of the peripherals have their own control stores.

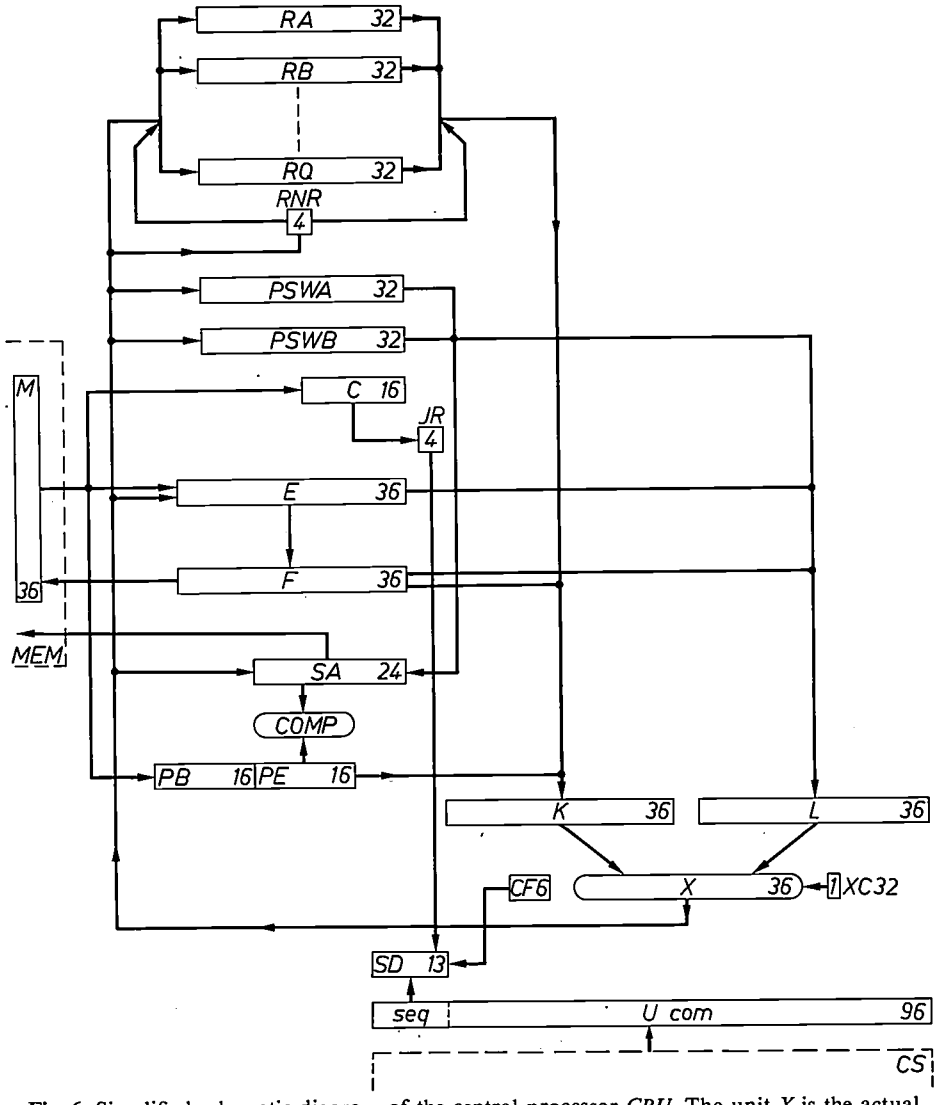


Fig. 6. Simplified schematic diagram of the central processor *CPU*. The unit *X* is the actual arithmetic unit. The number of bits contained in each register is shown. The working memory *M* is connected with the central processor via the buffer register *M*, and the control store *CS* is connected to *CPU* via the buffer register *U*.

The central processor

Fig. 6 shows a simplified schematic diagram of the central processor of a computer from the P1000 series. It contains a large number of registers, most of them having a capacity of 32 bits, an arithmetic unit *X* and the buffer registers *U* and *M* of the control store and the working memory *MEM*. The data is transferred between the various units along the data paths shown and is controlled by gates like those in fig. 1. These gates are not shown in the figure, and not all of the data paths are shown.

The actual arithmetic unit *X* is not a register, but merely a logic circuit. Its inputs are formed by the registers *K* and *L*, and the output can be connected to any one of the 16 registers *RA* . . . *RQ*. Indication of which register is selected is given by the contents of register *RNR*. The registers *RA* and *RB* (numbered 0 and 1) are used as accumulators, the others as index registers. The output of the arithmetic unit can also be connected to the registers *E*, *SA*, *PSWA*, *PSWB* and *RNR*.

The addresses in the working memory are selected from register *SA*, which must therefore be occupied by an address before reading or writing into the working memory. Data is transferred to and from the working memory via the buffer register *M*. The address parts of instructions fetched from the memory are put into register *E* and the operation parts in register *C*; numbers are transferred from *M* to *E*. Data is always fetched from the memory in whole words of 32 bits; if only 8 bits or half a word are required a *mask* of four bits is used to indicate which part of the word must be processed. Thus mask 0010 means that only the third group of eight bits is involved, 1100 indicates the first half of the word, and so on. The mask is derived from the operation code and the address.

The registers for the base address *PB* and final (end) address *PE* of a program and the comparison circuit *COMP* play a part in the address calculation. The registers *PSWA* and *PSWB* each contain one half of the program-status word, a combination of 64 bits that contains all the data of importance for the state of execution of the program being run. All these concepts will be discussed later with the instruction code.

The register *U* is the output buffer of the control store: the address selection in this store takes place from register *SD*, which can be filled from the sequence field of *U*, from the jump register *JR* or from a choice of no more than three of the six condition flipflops *CF*. The registers *JR* and *CF* play a part in the jump mechanism of the microprogram control, to be discussed later.

The arithmetic unit can operate in 16 different states (modes), according to the sort of operations that have to be performed, e.g. addition or subtraction, execution of the logic operations AND or OR with each bit, etc. The 'carry-in' required for working with negative numbers is stored in the flipflop *XC32*.

Timing

For efficient operation of the machine the working memory, the control store and the central processor must be well matched to one another. The computers of the P1000 family work with clock pulses at a period of 500 ns. These clock pulses are available in four phases relatively displaced by 125 ns (*fig. 7*), so that there is a clock pulse every 125 ns. The interval of 125 ns is called a pulse time; transferring data into a register takes less than one pulse time.

The cycle time of the working memory is 1 μs, which means that the contents of 10^6 addresses could be transferred to *CPU* in 1 s. The access time is half of this value, so that the data is available from the store within 500 ns. The control store is twice as fast: the cycle time is 500 ns and the data is available after 250 ns.

An operation cycle of *CPU* lasts for four pulse times, i.e. 500 ns. It starts by writing the data present at the output of the arithmetic unit into a register, e.g. the accumulator *RA* (one pulse time), and continues by filling the input registers *K* and *L* with new data (one pulse time). The arithmetic unit then takes two pulse times to produce a result.

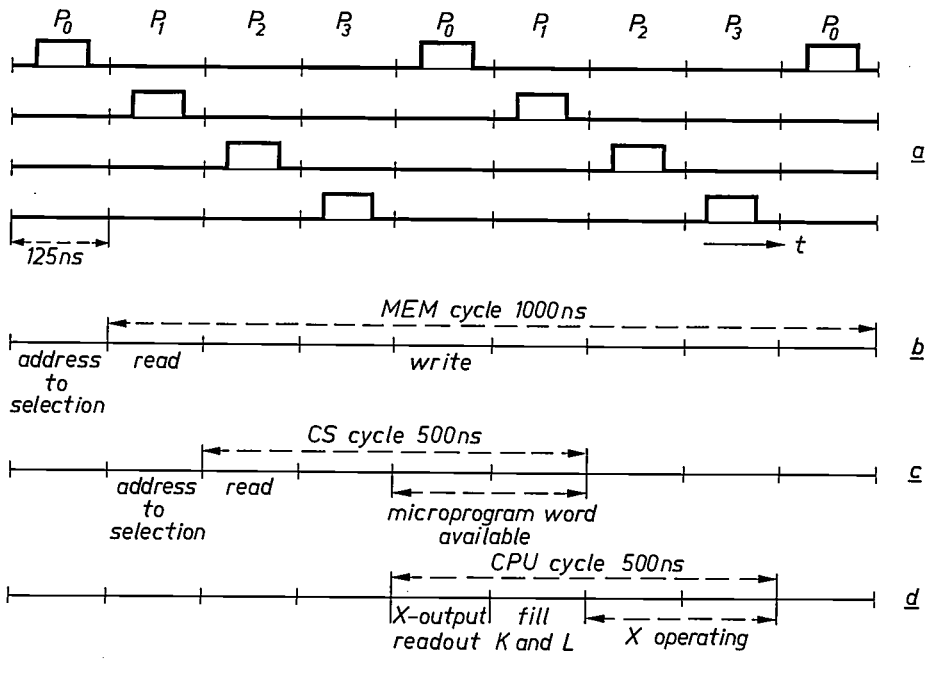


Fig. 7. Timing diagram of the operation of a P1000 computer. *a*) The four clock-pulse phases. *b*) The working-memory cycle. *c*) The control-store cycle. *d*) The cycle for the central processor.

The relative timings are shown in fig. 7. The intervals in which the various clock pulses appear are successively P_0, P_1, P_2 and P_3. During P_0 and P_1 the data of a microprogram word is always available; registers are then filled and commands for the working memory are supplied. During P_3 and P_4 the arithmetic unit has time to form its output and a new microprogram word is fetched from the control store. In our discussion of an example of a microprogram we shall make use of these relative time relationships.

The instruction code

The P1000 family of computers has about 150 instructions, which are the same for all models. We shall now say something about the format of these instructions with the aid of *fig. 8*. This figure shows the coding of the instruction 'add integer'. The intention of this instruction is that the contents (32 bits) of accumulator *RA* should be added to the contents (32 bits) of a memory location, and that the result should then be put back in the accumulator. Only integers are involved

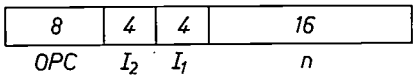


Fig. 8. The instruction code for 'add integer'. The first eight bits (*OPC*) give the operation code (here 01000100). I_1 and I_2 are two groups of four bits. These indicate index registers for modification of the address, or they indicate that the address must be taken indirectly. The last 16 bits form the address n of the number to be operated on. This address is expanded to 24 bits by modification, indirect addressing or by filling in noughts.

here. The first eight bits of the instruction give the operation code *OPC*; for the instruction 'add integer' this is the pattern 01000100. Two groups of four bits then follow, I_2 and I_1, which each indicate one of the 14 registers *RC* . . . *RQ* that must be used as index register and whose contents must be used to modify the address n from the instruction. It is therefore possible to modify twice in one instruction if necessary (first I_1, then I_2). The codes 0000 and 0001 (the numbers of the two accumulators, which cannot be used as index registers) here indicate respectively 'no modification' and 'the address n must be taken indirectly'.

The second modification (with I_2) will also work 'displaced'; in this case the contents of the index register are added to the address in a shifted position. Since the addressing in the working memory is for 8-bit groups, it is necessary to be able to increase the addresses by a number of units, but if for example it is desired to address in successive complete words (an 'integer' occupies a complete word, i.e. 4 groups of 8 bits), it is necessary to be able to increase addresses in steps of 4, which amounts to adding the number of units after shifting by two places.

The address n from the instruction is only 16 bits long; it is expanded to 24 bits (n') by modification or indirect addressing or just by filling in noughts. This length determines the actual number of addresses in the memory, the 'address area' (2^{24} or about 16 million 8-bit groups). The 24-bit address n', the relative address encountered earlier, is however not the absolute address, as we then saw, but still has to be added to the base address.

This method of address computation is used because in a modern computer it is desirable to be able to use multiprogramming, with a number of programs in the machine at different stages of execution. If an interrupt requires a program to be temporarily stopped, the housekeeping routine ensures that this program is changed for the one that has priority. The contents of the registers *RA* . . . *RQ* and also the program-status word are then stored and the corresponding data from the other program is substituted. If necessary the program in execution is put into the backing store and the new program is put into a free part of the working memory. The state of one of the programs is thus frozen while the other is continued after being frozen.

Each program contains a program number *PN* of four bits, which refers to tables in the working memory into which the housekeeping routine puts the new base address of the program at each change. This program number forms a subdivision of the program-status word; other subdivisions are 24 bits that function as an instruction counter *OT* (the state of this indicates the progress made with the program) and two *CC* bits. The *CC* bits give information about the result of the last instruction executed; for the adding instruction in our example it shows whether the result is equal to 0, smaller than 0, greater than 0 or whether there is an overflow i.e. the register is too small to contain the data. This information is important in the execution of a conditional jump instruction.

Since later on we shall take an example of a microprogram for the P1400 computer, we shall now look at the two ways in which the absolute address can be calculated in this computer.

1) The *program-base method PBM*. Here the complete program is put into the working memory, starting from a particular base address. This base address, the *program base PB*, must then be added to each address in the program (*fig. 9*). The program base is a word of 16 bits; the eight least-significant bits are missing. This means that the base address for locating a program in the memory is always a multiple of 256.

The other programs already in the working memory are protected by comparing, during the execution of the program, all the absolute addresses at the instant that they are in *SA* (see fig. 6) with the base and end addresses of the program (the start and finish of the program are then at *PB* and *PE* [4]). If the address in *SA* is outside the permitted area, access to the memory is blocked and this can be a reason for interrupting the execution of the program and starting another one. The comparison is made in the circuit *COMP*. *PB* and *PE* are filled at the start of the program. The desired data is found by way of the program number in the program table in the working memory.

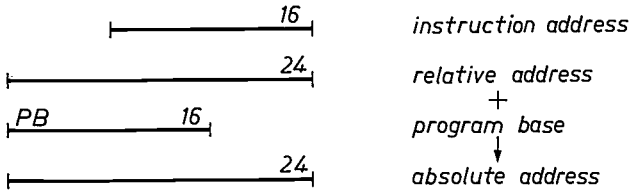


Fig. 9. Calculation of the absolute address by the program-base method *PBM*. The address from the instruction is expanded by modification or indirect addressing to an address of 24 bits, the relative address; adding the program base *PB* then gives the absolute address.

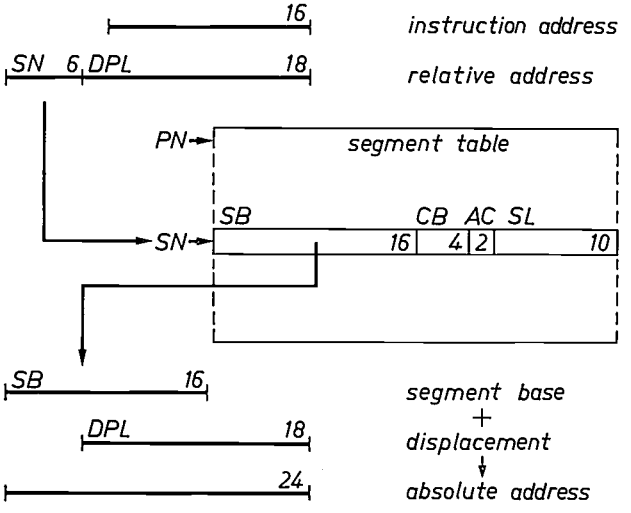


Fig. 10. The address calculation used in the segment-base method *SBM*. The instruction address, which is expanded to the relative address of 24 bits length after modification or indirect addressing, is divided into a segment number *SN* and a displacement *DPL* within that segment. For each program the memory contains a segment table, which is found by way of the program number *PN* (from the program-status word). The segment number refers to a word in this table that includes the segment base *SB* and the segment length *SL*. The absolute address is obtained by shifting the segment base eight places and adding the displacement.

2) The *segment-base method SBM*. In this method the program is divided into a number of segments, which can be distributed over the memory relatively independently of one another. Here the address is translated by adding to each address the base address of its segment, the *segment base SB* (16 bits). The segment data, such as the base, the segment length *SL* (for protection) and several control bits, is found in a segment table, one for each program. A segment is characterized by its segment number of 6 bits, so that the maximum number of segments that a program can have is 64. The input to the segment table is the program number; the input to a specific set of segment data is the segment number. The address calculation is as follows (*fig. 10*). The relative address, expanded to 24 bits, is split up into a segment number and another relative address, now within that segment, the 'displacement' *DPL* (18 bits). In the segment table, found via the program number, the 16 bits of the segment base and the segment length (here 10 bits, indicated in multiples of 256) are found in the word to which the segment number refers. This data is transferred to the registers *PB* and *PE*. The displacement is added to the segment base to give the absolute address. The maximum segment length is 2^{18} = about 260 000 8-bit groups.

The bits *AC* (access bits) indicate whether the segment can only be read, written in, or both. *CB* are control bits that show for example whether the segment is in the working memory or not.

One of the bits of the program-status word (the *PBM/SBM* bit) indicates which of the two methods of addressing is used in a particular program.

Microprogramming the P1400

We shall now look at the microprogramming of the P1400 computer. This has a control store that can contain 8192 words of 96 bits. When we were discussing timing relationships (fig. 7), we saw that a microprogram word for controlling the elementary operations is always available during the clock-pulse phases P_0 and P_1. Thus for an addition in the phase P_1, for example, the registers *K* and *L* are filled, during P_2 and P_3 the result is formed, and in the next phase P_0 the result is transferred to say *RA*. This series of operations can be represented graphically as in *fig. 11a*, in which time runs from top to bottom and the associated commands are shown. In the symbolism the letter *Y* represents an operator to be read as 'receives a copy of'.

The notation for the microprograms that we shall use here is derived from this figure: each microprogram word is represented by a vertical strip, divided into two parts corresponding to the pulse times P_0 and P_1. In the four regions *I*, *II*, *III* and *IV* thus formed (see fig. 11*b*) we indicate:

In *I* : the label of the word, i.e. the symbolic address; this label also shows the conditions under which the word is reached. The mode of the

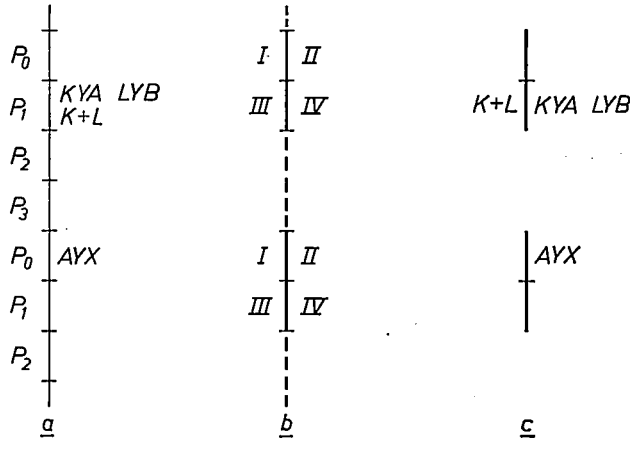


Fig. 11. *a*) Timing diagram for a short length of microprogram in which the contents of registers *A* and *B* are transferred to *K* and *L* respectively, the arithmetic unit performs an addition and the result is put back into *A*. Time runs from top to bottom. *b*) Notation for microprogram words. Each word is represented by a vertical strip representing the pulse phases P_0 and P_1 during which the word is available. The various labels and commands are written in the four regions *I*, *II*, *III* and *IV*. *c*) Notation in this convention for the length of microprogram in (*a*).

[4] For simplicity it is assumed here that *PE* contains the end of the program; in actual fact the program length is used.

working memory: *R* for read and *W* for write. The mask for the working memory, which indicates which 8-bit group is to be read from a 32-bit word.

In *II* : the commands that are active in phase P_0.

In *III* : the mode of the arithmetic unit.

In *IV* : the commands that are active in phase P_1, and also the label of the next microprogram word.

In fact the commands that are active for longer than one clock pulse are placed on the left of the strip. The pulse times P_2 and P_3, during which no commands can be given, are indicated only by a short space between the vertical lines of two successive microprogram words. Fig. 11*c* shows the piece of microprogram of fig. 11*a* in this convention.

In the general introduction to microprogram control we mentioned that a microprogram word is split into a command part and an address or sequence part. The command part is divided into different fields, which correspond to different condition signals. A complete discussion here would take us into too much detail. We shall therefore only look at the most interesting feature, the address determination by means of the sequence part.

The sequence mechanism

Normally each sequence part contains the address of the next microprogram word; we refer to this as 'absolute addressing'. Since the microprograms of many instructions have many parts in common, space in the control store can be saved by common use of these parts. This is one of the reasons for having an extensive mechanism of conditional jumps.

The method of addressing is indicated by the bit No. *13* of the microprogram word, the absolute/conditional or *A/C* bit. If this is 0 it indicates an absolute address and the 13 bits *0* to *12* form the new address (*fig. 12*). If the *A/C* bit is 1 it indicates a conditional address; bit *14* (the *J* bit) of the sequence field then determines the choice between the mechanisms of *fig. 13* ($J = 0$) or *fig. 14* ($J = 1$). In fig. 13 the first six bits from the old address are transferred unchanged to the new address. Four of the next seven bits are taken from the sequence field of the old word, while the remaining bits *a*, *b*, *c* and *d* are filled with the choice from three of the six condition flipflops *CF*. In fig. 13 it is assumed that the bits labelled a_1, a_2 and a_3 from the sequence field have the binary value $101 = 5$ so that they refer to CF_5, whose contents are transferred to bit *a* of the next address. The bits *b* and *c* of the next address are filled in a similar way. With this mechanism the microprogram can always make an 'eight-way jump'; the addresses to which jumps can be made are found in eight successive words. Depending on the state of the condition flipflops in use the microprogram can thus be continued in eight different ways. If the bits a_1, a_2 and a_3 have the contents 000 (or 001) then bit *a* of the new address is made 0 (or 1). Another way of putting it is that if the bits a_1 and a_2 of the sequence field contain the combination 00, bit *a* of the new address is transferred from bit a_3 of the sequence field. There is then one possible choice that has not been taken, giving a 'four-way jump'. If yet another possible choice is not taken the result is a 'two-way jump'.

The mechanism of fig. 14 can produce 16-way jumps. When the *J* bit is a 1, the first nine bits of the next address are taken from the sequence field and the last four bits are transferred from the jump register *JR*. Depending on the contents of this jump register the microprogram can now be continued in 16 different ways.

In the version that we shall give of a microprogram as the designer would write it, the various ways of making a jump follow from the method of labelling. The labels are given by combinations of three letters, such as AAA, AAB, AAC, ..., KAE, KAF, ..., etc. If it is desired to make a 16-way jump with the aid of the *JR* register, then Xs are chosen for the last two letters of the label, e.g. AXX. (There are therefore 16 possible addresses AXX, which could be designated by A0000 to A1111.)

If the other method for jumps is required (from fig. 13), then the three letters must be followed by the numbers of the flipflops that have to be inspected, e.g. KAE523. If a four-way jump is to be indicated, this is written as KAEx23, for example; a two-way jump could be given as KAExx3. If the value of the penultimate address bit is then to be made zero, we have KAEx03. *Fig. 15* gives a diagrammatic survey of a number of possible schemes with eight-way jumps.

An example of a microprogram

An example of a microprogram is given in *Table I*, which shows the microprogram for the instruction 'add integer' for the P1400. As the name indicates, whole numbers are to be added, and not numbers in say the floating-point representation (which would be much more complicated). The program has been somewhat simplified because it was not really practical to bring a number of the finer details into the explanation.

We shall first give a very general description of the microprogram. It consists of six words, some of which are however also used for a large category of other instructions. The six words do not necessarily appear in sequence in the microprogram memory; each word gives the label of the next one.

The example begins at the instant when the instruction has been fetched from the memory and put into

the registers C and E in the central processor (just how it got this far will be explained later in the discussion of this microprogram in relation to the next machine instruction); the execution of the instruction has not started at all as yet. At this instant checks are made to see if the program being run should perhaps be interrupted for something more urgent and to find out whether address modification or indirect addressing is desired with this instruction. In all these cases the microprogram is suspended after the first word, to be taken up again when these matters have been dealt with.

In the second microprogram word the first four bits of the operation code are investigated, and the last four are investigated in the third word. If is found that these bits refer to another machine instruction the microprogram is suspended; only if the operation code indicates the add-integer instruction is the fourth microprogram word executed. This word therefore controls the actual addition.

The fifth word controls the transfer of the sum to the accumulator, and establishes a number of characteristics of this result, for example whether it was zero, what sign it had and whether the register capacity was

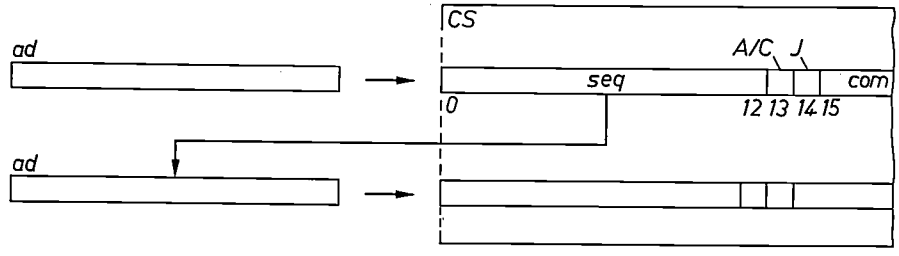


Fig. 12. The sequence mechanism in the microprogram of the P1000 computers. *CS* control store. Bits *0* to *12* address or sequence field of a microprogram word. Bit *13* *A/C* bit. Bit *14* *J* bit. Bits *15* to *95* command part. If the *A/C* bit is a 0, the 13 bits *0* to *12* form directly the address *ad* of the next microprogram word. If the *A/C* bit is a 1, then the next address is calculated as in fig. 13 or fig. 14.

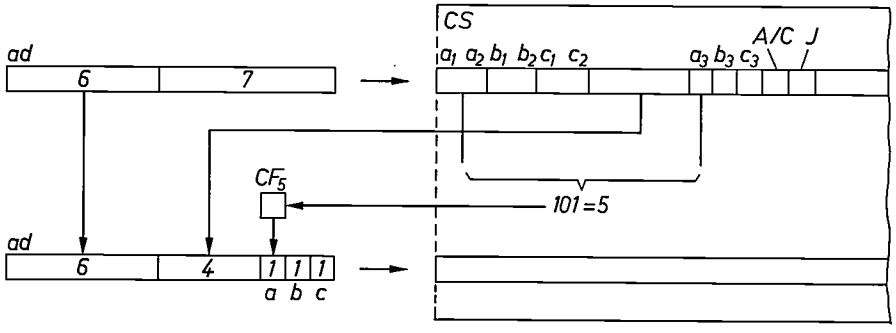


Fig. 13. Address determination in the microprogram if the *A/C* bit = 1 and the *J* bit = 0. The first six bits are taken from the old address, the next four bits are transferred from the sequence field of the microprogram word, the next three bits, labelled *a*, *b*, *c* are taken from three condition flipflops. In the example it is assumed that the combination of bits a_1, a_2 and a_3 from the sequence field has the binary value 101 (= 5). This refers to condition flipflop No. *5* and the contents of CF_5 are used for *a*. Bits *b* and *c* are filled in a similar way.

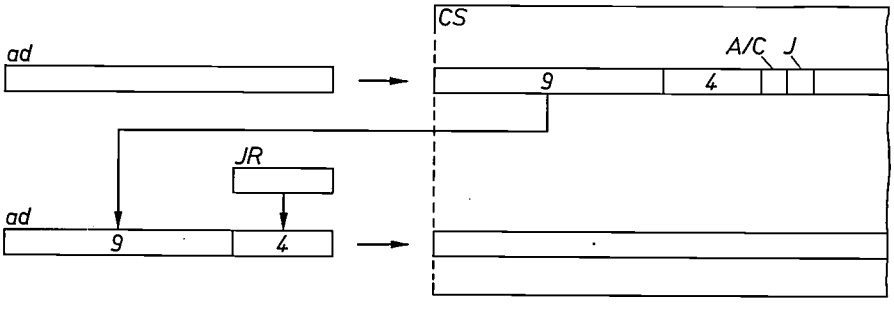


Fig. 14. Address determination in the microprogram if the *A/C* bit = 1 and the *J* bit = 1. In this case the first nine bits are taken from the sequence field and the remaining four bits from the jump register *JR*.

exceeded. Only in the case of such an overflow is the last microprogram word run. This word checks whether the overflow was expected or not. If it was unexpected the program may not be continued and a 'status switch' starts another program.

We shall now look at the example of Table I in detail.

The symbolic address of the first microprogram word, i.e. its label, is KAFx00. The conditional jumps at the end of the fifth and sixth words therefore jump to here if the contents of the condition flipflops CF_5 and CF_3 are zero. The label of the sixth microprogram word is KAFx01, where the microprogram continues after the fifth word if CF_5 is one. Since this sixth word checks capacity overflow, this overflow is evidently registered in CF_5.

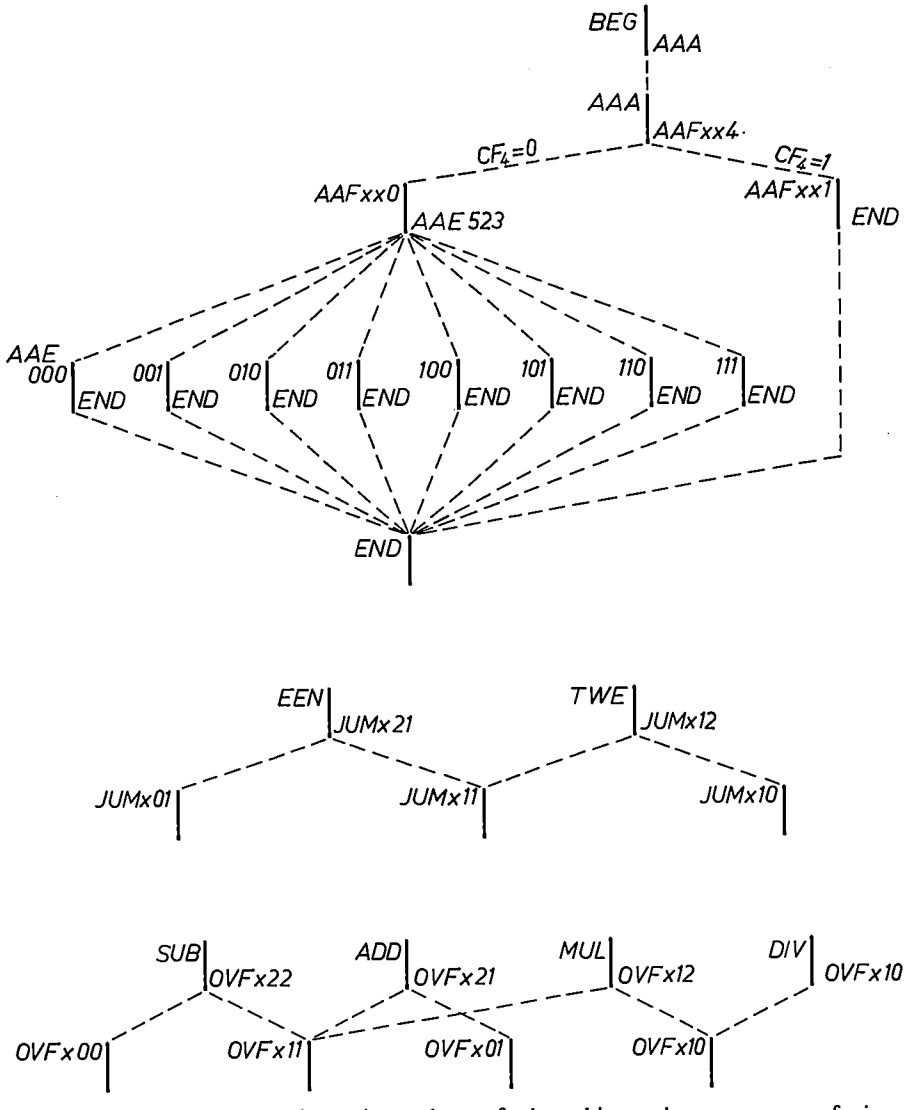
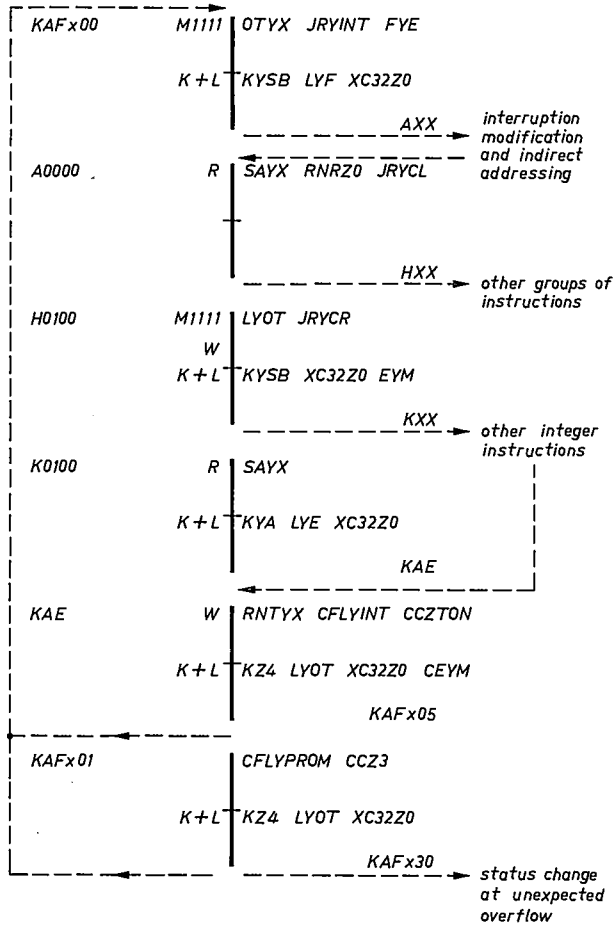


Fig. 15. Three diagrams showing various schemes for branching and reconvergence of microprograms with the use of two-way, four-way and eight-way jumps. The time again runs from top to bottom. Each vertical strip represents a period in which a particular microprogram word is active. Each word is characterized by its label (symbolic address), shown top left, and shows the label of the next word below right. Labels that indicate a conditional jump give the numbers of the condition flipflops that determine the address to which the jump is to be made.

During all these steps in the microprogram, preparations to fetch a following machine instruction are also encountered. Thus during the third microprogram word the address of the next instruction is calculated and during the fourth microprogram word the read/write cycle of the memory is started, which makes this new instruction available during the fifth word and enables it to be applied to the *C* and *E* registers. At the same time the instruction count in *OT* is raised. Everything is then ready for a new program cycle to begin.

In the first microprogram word the abbreviation OTYX indicates that a result is transferred from the arithmetic unit *X* to the *OT* register. This fixes the new value of the instruction count in *OT* (which is a part of the program-status word). INT is the indication for a number of conditions connected with various causes of an interrupt; JRYINT puts these into the *JR* register, so that in due course they determine the 16-way jump to label AXX. The address from the instruction is also transferred from *E* to *F* (FYE). All this takes place during phase P_0.

Table I. The simplified microprogram for the instruction 'add integer' in the Philips P1400 computer.



In phase P_1 preparations are encountered for the calculation of the absolute address of the operand in the add-integer instruction. The mode of the arithmetic unit is set to 'add' ($K + L$), the incoming carry to the arithmetic unit is set to 0 by the command XC32Z0 (for subtraction this would have been a 1; Z is an operator read as 'is filled with'). K is filled with the program or segment base by KYSB, and L is filled from F with the operand address from the instruction (LYF).

No new commands are given during the phase P_2 and P_3, as was stated in the discussion of fig. 7, but the arithmetic unit has the time to form the result; at the same time the new microprogram word, denoted by the conditional label AXX, is fetched from the control store. In the case of an interrupt, modification or indirect addressing the example is suspended by a jump to another word, but here we shall assume that this is not the case and that AXX (because of the contents of *JR*) refers to the second word in our example.

In phase P_0 of this word the absolute address now obtained is transferred to the address-selection register of the working memory (SAYX). Address protection operates here automatically, i.e. by logic, not microprogramming; if necessary access is blocked to the memory cycle, which starts in the next phase. During P_0 the register *RNR* is also set to 0 (RNRZ0), indicating the accumulator that will shortly receive the result of the add-integer instruction. The left-hand part of the operation code, CL, indicating the instruction group, is transferred to the jump register to initiate a different processing of various instruction groups (JRYCL).

During phase P_1 of this word several things happen that we shall for simplicity omit, and the memory cycle for the working memory starts, in which the number to be added will be fetched from the memory. The label HXX indicates a possible jump to other instructions, but in our case (CL = 0100) it just refers to the next word in the diagram (H0100).

During phase P_0 of this third word of the microprogram the mask M1111 tells us that four 8-bit groups are going to be processed simultaneously, while the memory cycle has progressed far enough for writing to begin (W). The number to be added should now be available in the memory buffer *M*. In this phase preparations are also encountered for fetching a new instruction, with the copying of the instruction counter into L as a prelude to the address calculation (LYOT). The right-hand part of the operation code, CR, is transferred to the jump register, to initiate a branching of other integer instructions (JRYCR).

In the phase that follows next the memory buffer is read (EYM), which puts the number into E, and the segment base is transferred to K (KYSB), so that the absolute address of the next instruction is obtained by adding ($K + L$); here again the command XC32Z0 provides the correct incoming carry. The label KXX in this case refers to the next word for the execution of the add-integer instruction (CR = 0100).

This fourth word begins with the transfer of the absolute address just calculated from the outputs of the arithmetic unit to the selection register of the working memory (SAYX); since the previous memory cycle is now ready, a read command is given for a new cycle (R); this will fetch the next instruction.

During phase P_1 of this microprogram word the accumulator contents *RA* and the number to be added are transferred to the inputs K and L of the arithmetic unit (KYA and LYE), and the incoming carry is set at 0 (XC32Z0), so that the addition can take place in the next two phases. This is in fact the actual execution of the instruction 'add integer'. The label KAE is absolute and refers to the next word. When that word is available, two phases later, the addition is also ready and the result can be transferred to the register denoted by *RNR* (RNTYX). In this example register 0 was indicated for this, i.e. the accumulator *RA*. At the same time the condition flipflops and the condition registers in *PSWB* are filled by CFLYINT and CCZTON; the only thing that affects the sequence here is that CF_5 indicates any overflow from the addition. The memory cycle for fetching the new instruction is by now so far advanced that this new instruction is in *M* and writing can begin (W). In the phase P_1 of this fifth word everything is made ready to increase the count of the instruction counter: the incoming carry is put at 0 (XC32Z0), *OT* is transferred to L (LYOT) and the number '4' to K (KZ4) — since 4 groups of 8 bits are being processed simultaneously the address must be increased by 4. CEYM sends part of the new instruction to the C register and the remainder to the E register; there is now a jump back to the start of the microprogram if $CF_5 = 0$, or forward to the next word if $CF_5 = 1$. In that case this next word deals with the overflow, and we shall say no more about it here.

A single machine cycle can in fact be considered to start with label KAE. The contents of this microprogram word are very nearly the same for all instructions, and are therefore common to a large number of instructions.

Processing of a microprogram design

As in the example given above, the microprogram is first written in a symbolic notation; finally, however, the address labels and all the commands must be converted into digital 0/1 patterns that can be put into the

control store. This conversion is almost completely automatic. It is done by punching the symbolic notation into cards or paper tape, and running special 'compile' programs that convert the symbolic commands into the 0/1 patterns of the various command fields, which will in fact later control the machine. The compile program also translates the address labels, in particular those for the jumps. While doing this, the program can ensure an optimum use of the locations in the control store: if for example only five of the jumps in an eight-way jump are used, the program can take note of this and use the three unused jump addresses again. On the other hand, if the microprogrammer wishes to keep the unused jumps open in case of later extensions he can record this; the three addresses are then kept in reserve.

The results of this automatic conversion can be made available in many ways. In the first place they will of course be available in the form required for the production of the control stores, but also in a variety of versions that can be read. Even at this stage statistical information can be obtained about the use and usefulness of particular microinstructions, and such information can be taken into account in considering other designs.

Another special program enables the microprogram designer to test his microprograms before they are built into a machine. This special program simulates the operation of the machine to be constructed by reading off microinstruction after microinstruction and making a close check of the operations that the actual machine would have performed. This makes it possible to track down any errors present and to correct them, which can save a great deal of time and trouble in testing the finished equipment.

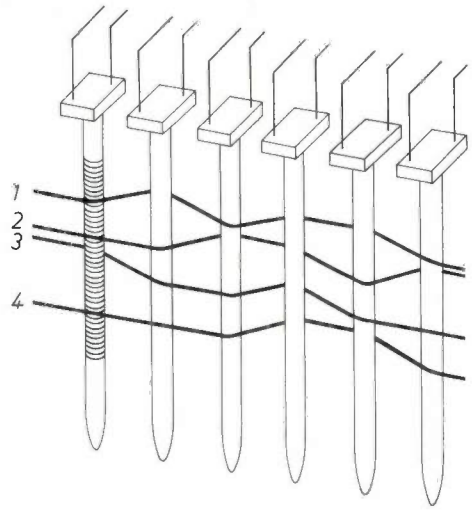


Fig. 16. Diagram to illustrate the principle of linear inductive coupling. This 'read-only' store consists of plastic rods, each closely wound by a winding that forms the secondary of a transformer. The primary winding is formed by a wire 'woven' between the rods. Each word corresponds to one such wire. A word is read from the store by putting a voltage pulse across its wire. Depending on whether the wire passes round the rods in a left-hand or right-hand sense positive or negative pulses are induced in the various secondary coils. This pattern of noughts and ones cannot be altered after the primary wire is threaded into place in manufacture.

Hardware

The control stores in the P1000 family are inductive stores, with linear inductive coupling. This means that a 0 is distinguished from a 1 by the sense in which a wire is passed around a tiny plastic rod, forming a kind of transformer (*fig. 16*). The rods are covered by a tightly wound coil that forms the secondary of the transformer. The 96 rods for the 96 bits of a microprogram word can clearly be seen in *fig. 17*. Corre-

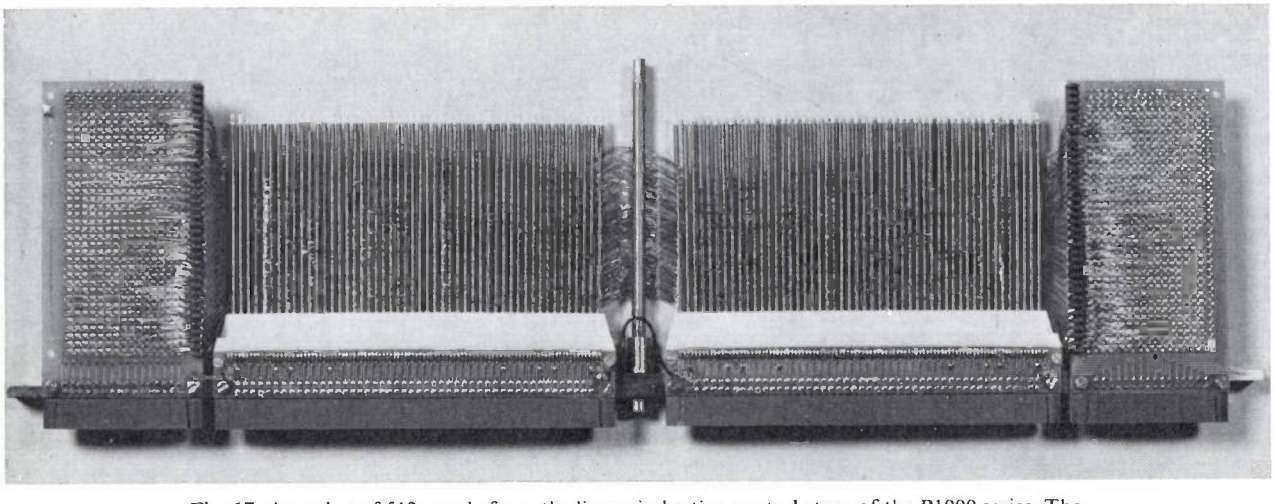


Fig. 17. A packet of 512 words from the linear-inductive control store of the P1000 series. The 96 rods and 512 word wires can be seen in the middle of the photograph. The panels at left and right contain the circuits that select a word wire and produce the 'read' pulse. The unit is folded round the central hinge in assembly.

sponding to 512 words, 512 different primary windings pass between the rods from left to right: the way in which they pass each rod determines the sense of the magnetic coupling. When a particular word is to be read from the store a pulse of current is passed through the associated primary winding by the circuits located to the left and right of the rods; the secondary windings then deliver the 96 bits of the word. The 0/1 pattern is built into the windings during manufacture and cannot be altered afterwards. These stores are known as 'read-only' memories [5].

The control store in the P1000 models is built up from packets of 512 words. The minimum memory capacity required differs for the various machine configurations, but can for example be 4 k. It can be extended beyond this minimum value in steps of 512 words. The maximum capacity in all models is 8 k words of 96 bits. The greater part of the control store is occupied by the microprograms of the instructions, but besides these there are also a number of microprograms for diagnostic functions. These are programs for checking the correct operation of a number of important parts of the machine; they are run by the machine at fixed times between the other programs.

Since not all of the users need to have all of the instructions, the machine instructions available are divided into groups, called 'instruction sets'. These include the *basic* set (in all models), the *decimal* set (for administrative applications), the *floating-point* set and the ALGOL set (for scientific calculations). A simple machine can always be extended by an extra instruction set by adding a few extra packets to the control store.

One last point: it is by no means essential for control stores to be of the 'read-only' type. Writable control stores are already in use in many new machines, and this greatly increases the flexibility. In the P1000 family the choice of control store was determined by the price/performance ratio of the basic electronic units available when the design was drawn up.

[5] See also for example page 362 of R. M. G. Wijnhoven, Hoofdgeheugens en besturingsgeheugens, Informatie **15**, 359-364, 1973 (No. 7/8).

Summary. In executing an instruction the central processor of a computer has to perform a large number of elementary operations, mostly transfers of data between registers. The control unit, which controls these operations, operates in older machines with the aid of a counter. Since address modification or indirect addressing can occur for an instruction, and in particular since interrupts and program changes can arise in multiprogramming, the number of operations is surprisingly large even for simple instructions. The control counter is consequently very complicated. This is why microprogram control, proposed by M. V. Wilkes as early as 1951, is now generally used. In this method the various elementary operations for each instruction are recorded in a number of words in a special control store. Each microprogram word consists of an address or sequence part that gives the address of the next word of the instruction, and a command part that indicates the operation. In executing an instruction these words must be read in the correct order, the commands decoded and the associated operations set under way. The control logic is no longer 'wired logic' as in counter control, but is determined by the contents of the control store. Microprogramming is therefore a much more flexible method, though not so fast.

After an introduction to several concepts such as modification, indirect addressing and multiprogramming a description is given of a number of general features of the computers of the P1000 family. These computers, the models P1075, P1100, P1175, P1200 and P1400, all have the same instruction repertoire, but different speeds. Descriptions are given of the central processor, the timing of the working memory and the control store (cycle times of 1 μs and 500 ns respectively), the instruction code and the two methods of address calculation: the program-base and segment-base methods. In the description of microprogram control particular attention is paid to the sequencing. This is because microprograms for different instructions can have chains of words in common in the control store, so that in executing an instruction jumps to another part of the store sometimes have to be made. Two-, four-, eight- and 16-way jumps can be made in the P1000 computers; the methods used for this differ from those of other manufacturers. The P1000 control stores are 'read-only' stores with linear inductive coupling. They are made up from packets of 512 words of 96 bits; the maximum capacity is 8 k words. A rather simplified microprogram for the instruction 'add integer' in the P1400 is given in the article as an example.

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.

M. Adriaansz: Time-resolved probe measurements in transient gas discharges.
J. Physics E **6**, 743-745, 1973 (No. 8). *E*

G. A. Allen: Calculations on the performance of gallium arsenide photocathodes.
Acta Electronica **16**, 229-236, 1973 (No. 3). *M*

G. Armand, Y. Lejay (both with Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette) **& J. B. Theeten:** Ondes de surface sur la face (001) d'un cristal cubique face centrée en présence d'une monocouche adsorbée.
Le Vide **28**, 94-96, 1973 (No. 164). *L*

V. Belevitch & C. Wellekens: Internal equalization in filters.
Int. J. Circuit Theory & Appl. **1**, 179-186, 1973 (No. 2). *B*

R. Bernard (Hôpital St. Pierre, Bruxelles), **W. Rey & H. Vainsel** (Hôp. St. Pierre, Br.): La surveillance continue du rythme cardiaque par ordinateur; méthode exploitant l'onde *P* intra-auriculaire.
Arch. Mal. Coeur **66**, 439-441, 1973 (No. 4). *B*

F. Berz: Carrier heating effects in junctions at very low currents.
Solid-State Electronics **16**, 1067-1071, 1973 (No. 9). *M*

H. Bex: A new YIG filter permits broadband tunable mode separation.
Proc. 1973 European Microwave Conf., Brussels, Vol. 2, paper B.9.4. *A*

R. Bleekrode & H. van Tongeren: Measured and calculated Cs excited-state densities in Cs-Ar low-pressure discharges.
J. appl. Phys. **44**, 1941-1942, 1973 (No. 4). *E*

J. Bloem: Trends in the chemical vapour deposition of silicon.
Semiconductor Silicon 1973, editors H. R. Huff & R. R. Burgess, pp. 180-190. *E*

J. Bloem: Band bending at a growing silicon surface.
Semiconductor Silicon 1973, editors H. R. Huff & R. R. Burgess, pp. 213-226. *E*

J. Bloem (Philips Semiconductor Development Laboratory, Nijmegen) **& J. C. Brice:** The segregation coefficient of zinc in tin.
J. Crystal Growth **20**, 53-56, 1973 (No. 1). *M*

K. Board: Thermal properties of annular and array geometry semiconductor devices on composite heat sinks.
Solid-State Electronics **16**, 1315-1320, 1973 (No. 12). *M*

J. H. den Boef: A matching circuit and bucking current stabilizer for a detector in an ESR spectrometer.
Rev. sci. Instr. **44**, 778, 1973 (No. 6). *E*

H. van den Boom & J. C. M. Henning: An orthorhombic chromium-center in the spinel $MgAl_2O_4$.
J. Phys. Chem. Solids **34**, 1211-1216, 1973 (No. 7). *E*

J. van den Boomgaard: Unidirectional solidification of a liquid into two solid phases in a ternary system.
Metallurg. Trans. **4**, 1485-1490, 1973 (No. 6). *E*

J. van den Boomgaard & A. M. J. G. van Run: The influence of primary precipitates on the tensile strength of unidirectionally solidified (Fe, Cr) - $(Cr, Fe)_7C_3$ *in-situ* grown composites containing 30 wt % Cr.
J. Mat. Sci. **8**, 1095-1100, 1973 (No. 8). *E*

H. Bouma (Institute for Perception Research, Eindhoven): Der Einfluß zweier Mydriatica auf die statischen Lichtreaktionen der menschlichen Pupille.
Die normale und die gestörte Pupillenbewegung, Symp. Bad Nauheim 1972, pp. 216-221; 1973.

P. W. J. M. Boumans, R. F. Rumphorst, L. Willemsen & F. J. de Boer: Solid state photodiode system matched to high-gain low-noise d.c. and lock-in amplifiers for use in multichannel emission spectrochemical analysis.
Spectrochim. Acta **28B**, 227-240, 1973 (No. 7). *E*

G.-A. Boutry: Brève histoire de la photoémission.
Acta Electronica **16**, 127-136, 1973 (No. 2). *L*

P. C. Brandon & T. N. van Boekel-Mol: Properties of purified malic enzyme in relation to Crassulacean acid metabolism.
Eur. J. Biochem. **35**, 62-69, 1973 (No. 1). *E*

P. B. Braun, J. Hornstra, C. Knobler (University of Leiden), **E. W. M. Rutten** (Univ. Leiden) **& C. Romers** (Univ. Leiden): The conformation of non-aromatic ring compounds, LXXVIII. The crystal and molecular structure of the 3,20-bis(ethylenedioxy) analogue of provitamin D.
Acta cryst. **B 29**, 463-469, 1973 (No. 3). *E*

R. Bridgen: Pulsed measurement of transistor-base charge against collector current.
Electronics Letters **9**, 366-367, 1973 (No. 16). *M*

J.-J. Brissot: Problems related to the growth of artificial calcite single crystals.
Proc. 1st European Electro-Optics Markets and Technology Conf., Geneva 1972, pp. 127-132; 1973. *L*

J.-J. Brissot, F. Desvignes (SODERN, Limeil-Brévannes) **& R. Martres:** Organic semiconductor bolometric target for infrared imaging tubes.
IEEE Trans. **ED-20**, 613-620, 1973 (No. 7). *L*

M. Brouha & K. H. J. Buschow: Pressure dependence of the Curie temperature of intermetallic compounds of iron and rare-earth elements, Th, and Zr.
J. appl. Phys. **44**, 1813-1816, 1973 (No. 4). *E*

M. Brouha & A. G. Rijnbeek: A reliable 'Teflon' cell with many electrical leads for pressures up to 40 kilobars.
Rev. sci. Instr. **44**, 852-854, 1973 (No. 7). *E*

T. M. Bruton & E. A. D. White (Imperial College, London): Measurements of solution properties and the growth of single crystals of lead tantalate $PbTa_2O_6$.
J. Crystal Growth **19**, 341-350, 1973 (No. 4). *M*

K. Bulthuis: Laser power and vibrational energy transfer in CO_2 lasers.
J. chem. Phys. **58**, 5786-5794, 1973 (No. 12). *E*

K. Bulthuis & G. J. Ponsen: Relaxation of the 10^00 lower laser level in CO_2.
Chem. Phys. Letters **21**, 415-417, 1973 (No. 2). *E*

K. H. J. Buschow: Magnetic properties of DyCd and related CsCl-type compounds.
J. appl. Phys. **44**, 1817-1820, 1973 (No. 4). *E*

K. H. J. Buschow: Composition and stability of $CaCu_5$-type compounds of yttrium with iron and cobalt.
J. less-common Met. **31**, 359-364, 1973 (No. 3). *E*

K. H. J. Buschow & A. R. Miedema: Thermal expansion of $ZrFe_2$ and some rare-earth iron compounds.
Solid State Comm. **13**, 367-370, 1973 (No. 3). *E*

H. J. Butterweck (Eindhoven University of Technology) **& W. Mecklenbräuker:** Constant-frequency synthesis of microwave networks.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper B.1.3. *E*

G. Chabrier (Faculté des Sciences, Dijon), **P. Dolizy, G. Eschard, J.-P. Goudonnet** (Fac. Sci., Dijon) **& P.-J. Vernier** (Fac. Sci., Dijon): Détermination des paramètres optiques et photoélectriques de couches trialcalines.
Acta Electronica **16**, 203-210, 1973 (No. 2). *L*

M. G. Collet & L. J. M. Esser: Charge transfer devices.
Festkörperprobleme **13**, 337-358, 1973. *E*

M. G. Collet & A. C. Vliegenthart: Calculations on potential and charge distributions in the peristaltic charge-coupled device.
Philips Res. Repts. **29**, 25-44, 1974 (No. 1). *E*

J. Cornelissen (Philips Glass Division, Eindhoven) **& J. A. Waterman:** Die Viskositäts-Temperatur-Beziehung von Flüssigkeiten, III. Über den Zusammenhang zwischen den Konstanten von Viskositäts-Temperatur-Gleichungen mit drei Konstanten.
Materialprüfung **15**, 131-133, 1973 (No. 4).

J. Cornet & D. Rossier: Nature of the band-edge electronic states in As-Te semiconducting glasses.
Phil. Mag. **27**, 1335-1358, 1973 (No. 6). *L*

J. H. N. Creyghton, P. R. Locher & K. H. J. Buschow: Nuclear magnetic resonance of ^{11}B at the three boron sites in rare-earth tetraborides.
Phys. Rev. B **7**, 4829-4843, 1973 (No. 11). *E*

M. Davio & J.-J. Quisquater: Rectangular universal iterative array.
Electronics Letters **9**, 485-486, 1973 (No. 21). *B*

P. Delsarte: An algebraic approach to the association schemes of coding theory.
Thesis, Louvain 1973. (Philips Res. Repts. Suppl. 1973, No. 10.) *B*

P. Delsarte & J.-J. Quisquater: Permutation cascades with normalized cells.
Information and Control **23**, 344-356, 1973 (No. 4). *B*

A. M. van Diepen, H. W. de Wijn (State University of Utrecht) **& K. H. J. Buschow:** Temperature dependence of the crystal-field-induced anisotropy in $SmFe_2$.
Phys. Rev. B **8**, 1125-1129, 1973 (No. 3). *E*

H. Dimigen: Ein verbessertes Ionenätzverfahren.
Elektronik-Industrie **4**, EP 105-106, 1973 (No. 9). *H*

R. J. Dolphin: The analysis of estrogenic steroids in urine by high-speed liquid chromatography.
J. Chromatography **83**, 421-430, 1973. *M*

J. H. J. van Dommelen & P. Vries: A glasslike carbon sample holder for a high temperature Guinier camera.
J. Physics E **6**, 582, 1973 (No. 6). *E*

W. F. Druyvesteyn, J. W. F. Dorleijn & P. J. Rijnierse: Analysis of a method for measuring the magnetocrystalline anisotropy of bubble materials.
J. appl. Phys. **44**, 2397-2400, 1973 (No. 5). *E*

F. C. Eversteyn: Chemical-reaction engineering in the semiconductor industry.
Philips Res. Repts. **29**, 45-66, 1974 (No. 1). *E*

F. C. Eversteyn & G. J. P. M. van den Heuvel: Method for determining the metallurgical layer thickness of epitaxially deposited silicon from SiH_4 down to 0.5 μm.
J. Electrochem. Soc. **120**, 699-701, 1973 (No. 5). *E*

E. Fabre: Characterization of defects in GaAs by photoluminescence measurements.
Luminescence of crystals, molecules, and solutions, Proc. Int. Conf., Leningrad 1972, editor F. Williams, pp. 439-443; 1973. *L*

C. T. Foxon, J. A. Harvey (University of Surrey) **& B. A. Joyce:** The evaporation of GaAs under equilibrium and non-equilibrium conditions using a modulated beam technique.
J. Phys. Chem. Solids **34**, 1693-1701, 1973 (No. 10). *M*

G. Frank & S. Garbe: Photoemission of GaAs in the reflection and transmission mode.
Acta Electronica **16**, 237-244, 1973 (No. 3). *A*

N. V. Franssen: Generatorsystemen voor elektronische muziekinstrumenten.
T. Ned. Elektronica- en Radiogen. **38**, 9-15, 1973 (No. 1). *E*

R. C. French: Speech scrambling and synchronization.
Thesis, Brighton Polytechnic 1973. (Philips Res. Repts. Suppl. 1973, No. 9.) *M*

G. Frens: An experiment concerning the dispersion forces between very small metal spheres.
Physics Letters **44A**, 208-210, 1973 (No. 3). *E*

Y. Genin: A new approach to the synthesis of stiffly stable linear multistep formulas.
IEEE Trans. **CT-20**, 352-360, 1973 (No. 4). *B*

J. J. Goedbloed & M. T. Vlaardingerbroek: Noise in IMPATT-diode oscillators at large signal levels.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper A.2.2. *E*

J. M. Goethals: Some combinatorial aspects of coding theory.
A survey of combinatorial theory, editors J. N. Srivastava *et al.*, North-Holland Publ. Co., Amsterdam 1973, pp. 189-208. *B*

G. G. P. van Gorkom, J. C. M. Henning & R. P. van Stapele: Optical spectra of Cr^{3+} pairs in the spinel $ZnGa_2O_4$.
Phys. Rev. B **8**, 955-973, 1973 (No. 3). *E*

R. A. Gough (University of Bradford) **& B. H. Newton:** An integrated wide-band varactor-tuned Gunn oscillator.
IEEE Trans. **ED-20**, 863-865, 1973 (No. 10). *M*

P. C. M. Gubbens (Interuniversitair Reactor Instituut, Delft) **& K. H. J. Buschow:** Magnetic phase transition in Tm_2Fe_{17}.
J. appl. Phys. **44**, 3739-3741, 1973 (No. 8). *E*

G. J. van Gurp: Diffusion-limited Si precipitation in evaporated Al/Si films.
J. appl. Phys. **44**, 2040-2050, 1973 (No. 5). *E*

H. J. M. de Haan, D. J. Lishman (Mullard Ltd., Mitcham, Surrey), **A. A. A. G. de Bruin** (Philips-Electrologica, Apeldoorn) **& W. P. Goes:** Word selection by integrated magnetic needles in the 10^7-bit THEMIS store.
IEEE Trans. **MAG-9**, 39-45, 1973 (No. 1). *E*

J. H. Haanstra & A. T. Vink: Localized vibrations in GaP doped with Mn or As.
J. Raman Spectr. **1**, 109-115, 1973 (No. 1). *E*

J. Haisma & W. T. Stacy: Interference fringes due to magnetic domains in $FeBO_3$.
J. appl. Phys. **44**, 3367-3369, 1973 (No. 7). *E*

P. Hansen & J.-P. Krumme: The 'compensation wall', a new type of 180° wall in gallium-substituted YIG.
AIP Conf. Proc. **10**, Part 1, 423, 1973. *H*

P. Hansen, J. Schuldt & W. Tolksdorf: Anisotropy and magnetostriction of iridium-substituted yttrium iron garnet.
Phys. Rev. B **8**, 4274-4287, 1973 (No. 9). *H*

H. Haug: Transmission probabilities of high-frequency phonons through a solid-He-II interface.
Physics Letters **45A**, 170-172, 1973 (No. 2). *E*

H. Haug: On size effects in helium II below 1 K.
J. low Temp. Phys. **12**, 479-490, 1973 (No. 5/6). *E*

N. Hazewindus, A. M. M. Otten & A. Petterson: Minimizing the effects of disturbing magnetic fields in a cyclotron axial injection system.
Nucl. Instr. Meth. **111**, 181-188, 1973 (No. 1). *E*

M. Helmig & C. G. Sluijter: Stroboscopische registratie met televisie.
T.F.F. (Toegepaste Fotografie en Film) 1973, No. 3, 7-9. *E*

J. H. C. van Heuven: Conduction and radiation losses in microstrips.
Proc. 1973 European Microwave Conf., Brussels, Vol. 2, paper B.7.4. *E*

B. Hill & K. P. Schmidt: A fast access holographic memory.
Proc. 1st European Electro-Optics Markets and Technology Conf., Geneva 1972, pp. 224-228; 1973. *H*

W. K. Hofker (Philips Research Labs., Dept. Amsterdam), **H. W. Werner, D. P. Oosthoek** (Philips Res. Labs. Amsterdam) **& H. A. M. de Grefte:** Profiles of boron implantations in silicon measured by secondary ion mass spectrometry.
Radiation Effects **17**, 83-90, 1973 (No. 1/2). *E*

E. P. Honig: Molecular constitution of X- and Y-type Langmuir-Blodgett films.
J. Colloid & Interface Sci. **43**, 66-72, 1973 (No. 1). *E*

S. van Houten & G. H. F. de Vries: D.C. gas discharge displays.
Proc. 1st European Electro-Optics Markets and Technology Conf., Geneva 1972, pp. 324-330; 1973. *E*

J.-P. Hurault: Problèmes théoriques en photoémission.
Acta Electronica **16**, 173-180, 1973 (No. 2). *L*

J.-P. Hurault: Some magnetic properties of liquid crystals.
AIP Conf. Proc. **10**, Part 2, 1459-1475, 1973. *L*

J.-P. Hurault: Static distortions of a cholesteric planar structure induced by magnetic or ac electric fields.
J. chem. Phys. **59**, 2068-2075, 1973 (No. 4). *L*

B. B. van Iperen: Influence of transverse instability on the efficiency of IMPATT diodes.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper A.8.4. *E*

J. Jackson: Infra-red gas analysers using solid-state devices.
Proc. 1st European Electro-Optics Markets and Technology Conf., Geneva 1972, pp. 17-22; 1973. *M*

M. Jatteau: Exemple d'instrument de thermographie infrarouge permettant d'aborder l'étude quantitative des phénomènes thermiques de surface.
Cahiers de la Thermique No. 3, série A, mars 1973, pp. V.3/1-28. *L*

W. H. de Jeu & J. van der Veen: Instabilities in electric fields of a nematic liquid crystal with large negative dielectric anisotropy.
Physics Letters **44A**, 277-278, 1973 (No. 4). *E*

D. Kasperkovitz: A new current-routing logic counter.
Solid-State Electronics **16**, 883-886, 1973 (No. 8). *E*

E. E. de Kluizenaar, G. E. Thomas & W. Klerks: Een roterende lineaire U.H.V.-doorvoer.
Ned. T. Vacuümtechniek **11**, 63, 1973 (No. 4). *E*

A. J. R. de Kock & P. G. T. Boonen: The investigation of microdefects in high-purity silicon crystals by means of lithium decoration.
J. appl. Phys. **44**, 2816-2828, 1973 (No. 6). *E*

A. J. R. de Kock, P. J. Roksnoer & P. G. T. Boonen: Microdefects in swirl-free silicon crystals.
Semiconductor Silicon 1973, editors H. R. Huff & R. R. Burgess, pp. 83-94. *E*

E. Kooi & J. A. Appels: Selective oxidation of silicon and its device applications.
Semiconductor Silicon 1973, editors H. R. Huff & R. R. Burgess, pp. 860-879. *E*

H. Köstlin & A. Atzei (European Space Research and Technology Centre, Noordwijk): Present state of the art in conductive coating technology.
Photon and particle interactions with surfaces in space, editor R. J. L. Grard, publ. Reidel, Dordrecht 1973, pp. 333-341. *A*

B. Kramer, A. Farrayre, L. Hollan, E. Constant (Faculté des Sciences de Lille) **& G. Salmer** (Fac. Sci. Lille): A 22 percent C.W. efficiency solid state microwave oscillator.
1972 IEEE-GMTT Int. Microwave Symp., Arlington Heights, pp. 187-189. *L*

E. Krätzig: Critical magnetic fields of gapless superconducting films.
Phys. Stat. sol. (a) **18**, K 65-67, 1973 (No. 2). *H*

G. Krekow & J. Schramm: Druckelektrodenanordnung für elektrostatische Aufzeichnungsverfahren.
Feinwerktechnik + Micronic **77**, 219-225, 1973 (No. 5). *H*

J.-P. Krumme & H. Dimigen: Ion-beam etching of groove patterns into garnet films.
IEEE Trans. **MAG-9**, 405-408, 1973 (No. 3). *H*

J.-P. Krumme & P. Hansen: The compensation bubble: A new type of magnetic bubble.
J. appl. Phys. **44**, 3805-3807, 1973 (No. 8). *H*

K. E. Kuijk: A precision reference voltage source.
IEEE J. **SC-8**, 222-226, 1973 (No. 3). *E*

K. E. Kuijk & H. Hagenbeuk: Voltage-controlled phase shift of triangular waves.
IEEE Trans. **IM-22**, 183-184, 1973 (No. 2). *E*

J. van Laar: The physical model of negative electron affinity photoemitters.
Acta Electronica **16**, 215-227, 1973 (No. 3). *E*

D. E. Lacklison, G. B. Scott, H. I. Ralph & J. L. Page: Garnets with high magnetooptic figures of merit in the visible region.
IEEE Trans. **MAG-9**, 457-460, 1973 (No. 3). *M*

M. Laguës: Caractérisation d'une couche adsorbée par ségrégation de surface au moyen de mesures de photoémission.
Acta Electronica **16**, 251-259, 1973 (No. 3). *L*

M. Laguës & J. L. Domange: Etude des cinétiques de ségrégation.
Le Vide **28**, 100-102, 1973 (No. 164). *L*

J. Lohstroh & M. Otala: An audio power amplifier for ultimate quality requirements.
Audio Engng. Soc., 44th Conv., Rotterdam 1973, pp. 1-20. *E*

F. A. Lootsma (Philips Information Systems and Automation, Eindhoven): Convergence rates of quadratic exterior penalty-function methods for solving constrained-minimization problems.
Philips Res. Repts. **29**, 1-12, 1974 (No. 1).

F. K. Lotgering & A. M. van Diepen: Valencies of manganese and iron ions in cubic ferrites as observed in paramagnetic Mössbauer spectra.
J. Phys. Chem. Solids **34**, 1369-1377, 1973 (No. 8). *E*

H. H. van Mal, K. H. J. Buschow & F. A. Kuijpers: Hydrogen absorption and magnetic properties of $LaCo_{5x}Ni_{5-5x}$ compounds.
J. less-common Met. **32**, 289-296, 1973 (No. 2). *E*

G. Marie & J. Donjon: Single-crystal ferroelectrics and their application in light-valve display devices.
Proc. IEEE **61**, 942-958, 1973 (No. 7). *L*

F. Meyer, E. E. de Kluizenaar & D. den Engelsen: Ellipsometric determination of the optical anisotropy of gallium selenide.
J. Opt. Soc. Amer. **63**, 529-532, 1973 (No. 5). *E*

F. Meyer & J. J. Vrakking: The adsorption of oxygen on a clean silicon surface.
Surface Sci. **38**, 275-281, 1973 (No. 1). *E*

F. Meyer & J. J. Vrakking: Measurement of the ionization cross section of the L_{23}-shell of chlorine using Auger electron spectroscopy.
Physics Letters **44A**, 511-512, 1973 (No. 7). *E*

A. R. Miedema: The electronegativity parameter for transition metals: Heat of formation and charge transfer in alloys.
J. less-common Met. **32**, 117-136, 1973 (No. 1). *E*

A. Mircea, J. Magarshack, P. Lesartre & M. Mautref: Noise studies on Gunn diodes.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper A.7.4. *L*

F. Möllers & R. Memming: Spectroelectrochemical studies of the oxidation of dimethyl-p-phenylenediamine by reflection spectroscopy at SnO_2-electrodes.
Ber. Bunsen-Ges. phys. Chemie **77**, 879-885, 1973 (No. 10/11). *H*

J. H. Mooij (Philips Division Elcoma, Eindhoven): Electrical conduction in concentrated disordered transition metal alloys.
Phys. Stat. sol. (a) **17**, 521-530, 1973 (No. 2).

A. E. Morgan & W. J. M. van Velzen: Characteristic energy loss and Auger electron spectra of GaP (110).
Surface Sci. **40**, 360-374, 1973 (No. 2). *H, E*

G. Mörtl, W. Zednicek (both with Österr.-Amerik. Magnesit A.G., Radenthein, Österreich), **A. Odding & S. C. Rademaker** (both with Philips Glass Division, Eindhoven): Verhalten basischer Steine im Gitterwerk von Bleiglaswannen.
Glastechn. Ber. **46**, 141-147, 1973 (No. 7).

B. J. Mulder: Proto-djurleite, a metastable form of cuprous sulphide.
Kristall und Technik **8**, 825-832, 1973 (No. 7). *E*

K. H. Nicholas & R. A. Ford: Implanted resistors with properties enhanced by damage.
Proc. IEEE Int. Electron Devices Meeting, Washington 1973, pp. 51-53. *M*

H. Nosrati: A modified Butcher formula for integration of stiff systems of ordinary differential equations.
Math. of Computation **27**, 267-272, 1973 (No. 122). *B*

L. J. van der Pauw: Diffraction of a Bleustein-Gulyaev wave by a conductive semi-infinite surface layer.
J. Acoust. Soc. Amer. **53**, 1107-1115, 1973 (No. 4). *E*

R. I. Pedroso & G. A. Domoto (Columbia University, New York): Exact solution by perturbation method for planar solidification of a saturated liquid with convection at the wall.
Int. J. Heat & Mass Transfer **16**, 1816-1819, 1973 (No. 9). *N*

J. G. J. Peelen: Invloed van de microstructuur op de optische eigenschappen van keramische materialen.
Klei en Keramiek **23**, 170-177, 1973. *E*

R. Pepperl: Die digitale Laserstrahlablenkung und ihre Anwendungen.
Phys. Blätter **29**, 352-361, 1973 (No. 8). *H*

G. Piétri: The impact of negative electron affinity on vacuum tube technology.
Acta Electronica **16**, 267-271, 1973 (No. 3). *(Also in French, pp. 261-265.)* *L*

L. G. Pittaway: Modulation of the desorbed ion current in Bayard-Alpert gauges.
J. Vac. Sci. Technol. **10**, 507-512, 1973 (No. 4). *M*

R. J. van de Plassche: IC-elektronica toegepast in analoge bouwblokken.
T. Ned. Elektronica- en Radiogen. **38**, 47-56, 1973 (No. 2/3). *E*

J. J. A. Ploos van Amstel & E. Kooi: Porous carbides as evaporation sources for vacuum deposition of metal and semiconductor layers.
J. Electrochem. Soc. **120**, 840-843, 1973 (No. 6). *E*

I. Pockrand & J. Verweel: Observation of a new domain configuration in polycrystalline FeSi films.
Appl. Phys. Letters **23**, 276-278, 1973 (No. 5). *H*

J. Polman: Elektronentemperatuur-relaxatie in een zwak-geïoniseerd plasma.
Ned. T. Natuurk. **39**, 134-135, 1973 (No. 9). *E*

A. Rabier & R. Spitalnik: Diode characterization and circuit optimization for transferred electron amplifiers.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper A.6.1. *L*

G. Rau & J. Vredenbregt (Institute for Perception Research, Eindhoven): EMG-force relationship during voluntary static contractions (m. biceps).
Medicine and Sport **8**: Biomechanics III, 270-274, 1973.

H. Rau, T. R. N. Kutty & J. R. F. Guedes de Carvalho: Thermodynamics of sulphur vapour.
J. chem. Thermodyn. **5**, 833-844, 1973 (No. 6). *A*

W. J. J. Rey: Robust estimates of quantiles, location and scale in time series.
Philips Res. Repts. **29**, 67-92, 1974 (No. 1). *B*

J.-C. Richard: Photoémission du silicium.
Acta Electronica **16**, 245-250, 1973 (No. 3). *L*

E. D. Roberts: The preparation and properties of a polysiloxane electron resist.
J. Electrochem. Soc. **120**, 1716-1721, 1973 (No. 12). *M*

T. E. Rozzi: Network analysis of strongly coupled transverse apertures in waveguide.
Int. J. Circuit Theory & Appl. **1**, 161-178, 1973 (No. 2). *E*

T. E. Rozzi & J. H. C. van Heuven: Optimisation of microwave circuits by means of algebraic invariants.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper B.1.1. *E*

T. E. Rozzi & W. F. G. Mecklenbräuker: Field and network analysis of waveguide discontinuities.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper B.1.2. *E*

L. J. van Ruyven, H. M. Eijkman (both with Philips Semiconductor Development Laboratory, Nijmegen) **& J. J. K. Reinders** (Philips Information Systems and Automation, Eindhoven): Thin epitaxial layer thickness measurement by computer approximation of the interference pattern.
Semiconductor Silicon 1973, editors H. R. Huff & R. R. Burgess, pp. 616-623.

G. Salmer, I. Doumbia, B. Carnez (all with Université de Lille) **& A. Mircea:** 'Locally tuned' reflection type IMPATT diode amplifier.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper A.6.5. *L*

K. H. Sarges: The influence of a free surface on nuclear spin waves.
Phys. Stat. sol. (b) **59**, 403-409, 1973 (No. 2). *H*

C. J. Schoot, J. J. Ponjee, H. T. van Dam, R. A. van Doorn & P. T. Bolwijn: New electrochromic memory display.
Appl. Phys. Letters **23**, 64-65, 1973 (No. 2). *E*

M. F. H. Schuurmans & W. van Haeringen: The bound particle spatial density distribution in Appelbaum-Kondo's trial ground state of Kondo's Hamiltonian.
Solid State Comm. **13**, 171-173, 1973 (No. 2). *E*

J. Smith: Dc gas-discharge display panels with internal memory.
IEEE Trans. **ED-20**, 1103-1108, 1973 (No. 11). *M*

J. L. Sommerdijk & A. Bril: Visible luminescence of Yb^{3+},Er^{3+} under IR excitation.
Luminescence of crystals, molecules, and solutions, Proc. Int. Conf., Leningrad 1972, editor F. Williams, pp. 86-91; 1973. *E*

J. L. Sommerdijk, A. Bril, J. A. de Poorter & R. E. Breemer: Fluorescence decay of Yb^{3+},Er^{3+}-doped compounds, Part II. Cathode-ray excitation.
Philips Res. Repts. **29**, 13-24, 1974 (No. 1). *E*

R. Spitalnik, A. Rabier & J. Magarshack: Efecto del perfil de concentración en el comportamiento del amplificador a transferencia de electrones.
Electrónica y Fis. apl. **16**, 214-215, 1973 (No. 2). *L*

W. T. Stacy, M. M. Janssen, J. M. Robertson & M. J. G. van Hout: Dependence of the uniaxial magnetic anisotropy on the misfit strain in Gd,Ga:YIG LPE films.
AIP Conf. Proc. **10**, Part 1, 314-318, 1973. *E*

J.-B. Theeten: Contribution à l'étude des vibrations atomiques de surfaces propres ou recouvertes de monocouches ordonnées: mesures par diffraction d'électrons lents entre 20 K et 500 K interprétées dans des modèles simples.
Thesis, Paris-Sud 1973. (Philips Res. Repts. Suppl. 1973, No. 8.) *L*

H. Tjassens: Circuit analysis of a stable and low noise IMPATT oscillator for X-band.
Proc. 1973 European Microwave Conf., Brussels, Vol. 1, paper A.1.2. *E*

C. van Trigt: Asymptotic solutions of integral equations with a convolution kernel, I.
J. math. Phys. **14**, 863-873, 1973 (No. 7). *E*

C. van Trigt & J. B. van Laren: On radiative transfer in gas discharges.
J. Physics D **6**, 1247-1252, 1973 (No. 10). *E*

N. C. de Troye: Large scale integration. *(In Dutch.)*
Informatie **15**, 355-358, 1973 (No. 7/8). *E*

C. H. F. Velzel: Image contrast and efficiency of nonlinearly recorded holograms of diffusely reflecting objects.
Optica Acta **20**, 585-606, 1973 (No. 8). *E*

W. J. M. van Velzen & A. E. Morgan: Chemisorption on gallium phosphide surfaces.
Surface Sci. **39**, 255-259, 1973 (No. 1). *E, H*

J. P. M. Verbunt: Simple optical interference method for the inspection of solid surfaces.
Appl. Optics **12**, 1839-1840, 1973 (No. 8). *E*

L. Verhoeven: More aspects of quantisation noise associated with digital coding of colour-television signals.
Electronics Letters **9**, 69-70, 1973 (No. 3). *E*

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven): Luminescence of Mn^{2+} in $SrGa_{12}O_{19}$, $LaMgGa_{11}O_{19}$, and $BaGa_{12}O_{19}$.
J. Solid State Chem. **7**, 468-473, 1973 (No. 4).

J. F. Verwey: Nonavalanche injection of hot carriers into SiO_2.
J. appl. Phys. **44**, 2681-2687, 1973 (No. 6). *E*

J. Visser: Mass spectrometric analysis of the sputter gas atmosphere without pressure reduction system.
J. Vac. Sci. Technol. **10**, 464-471, 1973 (No. 3). *E*

J. Visser & J. J. A. Boereboom: An improved cryosorption pump design.
Ned. T. Vacuümtechniek **11**, 26-27, 1973 (No. 2). *E*

J. Visser & J. J. Scheer: The influence of the gas distribution on the performance of cryosorption pumps.
Ned. T. Vacuümtechniek **11**, 17-25, 1973 (No. 2). *E*

J. Vredenbregt & G. Rau (Institute for Perception Research, Eindhoven): Muscle coordination in simple movements.
Medicine and Sport **8**: Biomechanics III, 239-242, 1973.

L. Vriens: Light scattering from weakly ionized nonhomogeneous plasmas.
Phys. Rev. Letters **30**, 585-588, 1973 (No. 13). *E*

L. Vriens: Energy loss of charged particles in a plasma.
Phys. Rev. A **8**, 332-339, 1973 (No. 1). *E*

D. Washington: Improving the resolution of high gain channel plate arrays for particle and photon counting.
Nucl. Instr. Meth. **111**, 573-576, 1973 (No. 3). *M*

J. H. Waszink: A non-equilibrium calculation on an optically thick sodium discharge.
J. Physics D **6**, 1000-1006, 1973 (No. 8). *E*

W. F. van der Weg (Philips Research Labs., Dept. Amsterdam), **H. E. Roosendaal & W. H. Kool** (both with F.O.M.-Instituut voor Atoom- en Molecuulfysica, Amsterdam): High resolution measurements of energy spectra of protons scattered from silicon crystals in the case of planar channelling.
Radiation Effects **17**, 91-97, 1973 (No. 1/2).

W. F. van der Weg (Philips Research Labs., Dept. Amsterdam), **W. H. Kool** (F.O.M.-Instituut voor Atoom- en Molecuulfysica, Amsterdam), **H. E. Roosendaal** (F.O.M.-Inst. A. & M., Amsterdam) & **F. W. Saris** (F.O.M.-Inst. A. & M., Amsterdam): Silicon surface studies by means of proton backscattering and proton induced x-ray emission.
Radiation Effects **17**, 245-252, 1973 (No. 3/4).

K. Weiss & H. Haug: Bose condensation and superfluid hydrodynamics.
Cooperative Phenomena, editors H. Haken & M. Wagner, publ. Springer, Berlin 1973, pp. 219-235. *E*

H. W. Werner, H. A. M. de Grefte & J. van den Berg: Application of characteristic secondary ion mass spectra to a depth analysis of copper oxide on copper.
Radiation Effects **18**, 269-273, 1973 (No. 3/4). *E*

J. te Winkel: Extended charge-control model for bipolar transistors.
IEEE Trans. **ED-20**, 389-394, 1973 (No. 4). *E*

A. W. Witmer: Spektralanalysen. Röntgenfluoreszenzanalyse.
Techn. Rdsch. **64**, No. 6, 25-27, 1972. *E*

A. W. Witmer: Spektralanalysen. Massenspektrometrische Analyse anorganischer Feststoffe.
Techn. Rdsch. **65**, No. 9, 37-39, 1973. *E*

S. Wittekoek, J. M. Robertson, T. J. A. Popma & P. F. Bongers: Faraday rotation and optical absorption of epitaxial films of $Y_{3-x}Bi_xFe_5O_{12}$.
AIP Conf. Proc. **10**, Part 2, 1418-1422, 1973. *E*

W. K. Zwicker & S. K. Kurtz: Anisotropic etching of silicon using electrochemical displacement reactions.
Semiconductor Silicon 1973, editors H. R. Huff & R. R. Burgess, pp. 315-326. *N*

Contents of Philips Telecommunication Review **32**, No. 1, 1974:

Contents of Electronic Applications Bulletin **32**, No. 2, 1974:

Contents of Mullard Technical Communications **13**, No. 122, 1974:

Contents of Valvo Berichte **17**, No. 4, 1973:

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 7

Small electric motors II

This second issue on small electric motors contains four articles. Two of them are about motor designs, one article deals with the interesting problems that can arise when it is desired to control the speed of a capacitor motor with thyristors, while the first article in this issue describes the measurement of motor torque through the measurement of the magnetic field in the air gap. One of the two motors discussed is an induction motor for very high speeds, and the other is a thyristor-controlled d.c. motor that will operate at speeds suitable for a washing machine and spin drier. This motor is used in a Philips washing machine now on the market.

Torque measurements on induction motors using Hall generators or measuring windings

E. M. H. Kamerbeek

Introduction

In the development, testing and control of electric motors, and in research and teaching, it is frequently desirable to be able to measure and record the torque of an electric motor as a function of speed or time.

If T_e is the electromagnetic torque exerted on the rotor, then for a constant angular velocity the torque delivered will be equal to T_e less the frictional torque. Where the angular velocity is not constant (during transients) there is a third torque to be taken into account, which is the accelerating or retarding torque exerted on the rotor. In this article a method is described that is suitable for direct measurement or recording of the torque T_e, irrespective of the operating conditions. In this method the radial component of the magnetic field in the air gap between stator and rotor is measured. This is done by introducing a number of field probes (i.e. Hall generators or 'measuring windings') into the air gap of the motor under test. The number of Hall generators depends closely on the type of motor and on the required measuring accuracy, whereas the number of measuring windings depends only on the type of motor. The auxiliary equipment needed is mainly electronic. Other methods of torque measurement are less suitable or completely unsuitable for measuring and recording transient effects since their speed of response is too slow. They also require elaborate, usually expensive, mechanical devices such as a movable stator-housing or a torsion sensor mounted on the shaft coupling the motor to the load.

For d.c. motors it is best to use Hall generators, whereas for a.c. motors Hall generators and measuring windings are both effective, though our investigations have shown that measuring windings give the best results [1].

The experimental results described in this article refer only to a number of experiments carried out on a three-phase induction motor of the wound-rotor type.

Electromagnetic forces and torques acting on current conductors and magnetized material

In an electric motor the forces and torques exerted on the *current conductors* as a result of the magnetic field can be derived from the vector product $\boldsymbol{J} \times \boldsymbol{B}$ inside the conducting material. Here J is the conduction current density and $\boldsymbol{B}$ the magnetic flux density; the vector product is referred to as the force density.

The *magnetized material* may be treated in magnetic terms as the equivalent of a rotational current distribution with the density J_m. This current density is con-

Dr Ir E. M. H. Kamerbeek is with Philips Research Laboratories, Eindhoven.

[1] A detailed treatment has been given in: E. M. H. Kamerbeek, On the theoretical and experimental determination of the electromagnetic torque in electrical machines, thesis, Eindhoven 1970 (also published as Philips Research Repts. Suppl. 1970, No. 4).

nected with the magnetization M by the vector relationship $J_m = \nabla \times M$. If we confine ourselves to a calculation of total forces and torques, and assume that the parts of the material considered are surrounded by air or free space, we can use the fictitious force density $J_m \times B$ for the magnetized material. (It is called 'fictitious' to distinguish it from the real force density that produces the mechanical stresses in the material.) The model in which distributed rotational currents represent the magnetized material considerably simplifies the further treatment of the magnetic forces.

The electromagnetic force

When a body (of volume V) consisting of current-conducting and magnetized material is placed in a magnetic field, the electromagnetic force F_e acting on that body is given by:

$$F_e = \int_V (J + J_m) \times B \, dV. \qquad (1)$$

If we surround the volume V with a surface A as in

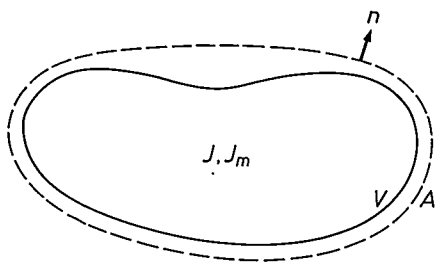


Fig. 1. A body of volume V contains current-conducting material (current density J) and magnetized material (represented by rotational currents with a density J_m). The volume integral of the electromagnetic forces can be reduced to a surface integral over a closed surface A (with normal vector n) outside the body.

fig. 1, it follows from the theory of the electromagnetic field that the volume integral (1) is equivalent to a surface integral over A:

$$F_e = \frac{1}{\mu_0} \int_A \{(B.n)B - \tfrac{1}{2} nB^2\} \, dA. \qquad (2a)$$

In this equation μ_0 is the magnetic permeability of free space and n is the outward-directed unit vector normal to the closed surface A.

Outside the body the magnetic flux density B is equal to $\mu_0 H$, where H is the magnetic field-strength. We can therefore write equation (2a) in the form:

$$F_e = \int_A \{(\mu_0 H.n)H - \tfrac{1}{2} \mu_0 n H^2\} \, dA. \qquad (2b)$$

The expression (2b) is a valuable aid in the calculation of electromagnetic forces. The information wanted is the field at the position of the surface A. The integrand of the surface integral (2b) has the dimension of a mechanical stress. In electromagnetic field theory it is known as the Maxwell or electromagnetic stress.

As an example of the ease with which certain problems can be handled with the aid of equation (2b) we can consider the paradoxical observation that while conductors embedded in slots in the rotor are the cause of a tangential force acting on the rotor, they themselves are subjected to this force only to a very slight extent. The situation is illustrated in *fig. 2*. In fig. 2*a* a rotor conductor lies on the circumference of the rotor R (developed into a plane). In the uniform magnetic field (with flux density B_s) under the stator pole S the rotor conductor is subjected to a force per metre length of $F_{e,x} = B_s I_r$ in the x-direction when a current I_r flows through this conductor.

If the rotor conductor is in a slot (fig. 2*b*) it is subjected to a much weaker magnetic field, because the lines of force curve sideways into the upper part of the slot and enter the rotor iron. Nevertheless, experiments show that the rotor as a whole is subjected to the same force $F_{e,x}$ as in fig. 2*a*; apparently a large proportion of the force now acts on the rotor iron and the remaining part on the conductor. Expression (2b) can be used for demonstrating theoretically that the force is of equal magnitude in both cases.

To do this the surface A, which encloses both the rotor iron and the rotor conductor, is made to have the special shape shown in fig. 2*b*. We further assume that the magnetic permeability of the stator and rotor iron is very high ($\mu_r > 1000$). In that case the lines of force are virtually perpendicular to the iron surfaces. The Maxwell stress $\mu_0(H.n)H - (\mu_0/2)nH^2$ then only has an x-component in the elements of the surface A that are vertical in the figure. At these surface elements the field in the air gap can be regarded as uniform. On the left this field has a value H_a and on the right a value H_b. Using relation (2b) we then find:

$$F_{e,x} = \frac{\mu_0}{2}(H_a^2 - H_b^2)\delta = \frac{\mu_0 \delta}{2}(H_a - H_b)(H_a + H_b), \qquad (3)$$

where δ is the air gap.

The two factors in (3) can be separately calculated. To calculate the sum of H_a and H_b we note that the fields H_a and H_b have components H_{sa} and H_{sb} respectively originating from the stator field H_s and components H_{ra} and H_{rb} originating from the rotor field H_r. The lines of force of the field H_s are represented in *fig. 3a*, and those of H_r in fig. 3*b*. By virtue of existing symmetries we have: $H_{sb} = H_{sa}$ and $H_{rb} = -H_{ra}$, and therefore:

$$H_a + H_b = 2H_{sa}. \qquad (4)$$

The difference between H_a and H_b is calculated with the aid of the contour integral $\oint H.ds$ along the closed path C_1 shown in fig. 2*b*. According to the theory of the electromagnetic field, the value of this integral is equal to the current I_r contained by the contour. Since H is negligibly small in the bulk of the stator and rotor iron, we can write:

$$(H_a - H_b)\delta = I_r. \qquad (5)$$

Substitution of (4) and (5) in (3) gives:

$$F_{e,x} = \mu_0 H_{sa} I_r. \qquad (6)$$

Thus, the x-component of the force acting on the rotor conductor plus the rotor iron is indeed equal to the force to which the conductor would be subjected in the configuration shown in fig. 2*a*, provided that $B_s = \mu_0 H_{sa}$. If the slot is narrow compared with the stator pole and the points a and b are at a sufficient distance from the slot, then $B_s = \mu_0 H_{sa}$ to a very good approximation.

The electromagnetic torque

Using the Maxwell stress we now seek an expression for the torque generated in an electric motor, in terms of parameters that can be measured while the motor is running.

Let us consider an induction motor with a cylindrical stator bore. The stator windings are contained in slots cut into the stator (*fig. 4*). We shall assume that the stator core possesses ideal magnetic and electrical properties, in other words that it has a very high magnetic permeability ($\mu_r \to \infty$) and negligible iron losses. In our further calculations we shall use the cylindrical coordinates r, ϕ and z.

It follows from the foregoing that the electromagnetic torque acting on the rotor can be found from the surface integral

$$T_e = \boldsymbol{i}_z \cdot \int_A r \times \{\mu_0(\boldsymbol{H}.\boldsymbol{n})\boldsymbol{H} - \tfrac{1}{2}\mu_0 n H^2\}\, dA, \quad (7)$$

provided that the surface A completely enclosing the rotor passes through air; $\boldsymbol{i}_z$ is the unit vector in the direction of the z-axis (which is at the same time the axis of rotation). For A we take the cylinder of revolution with radius a as shown in *fig. 5*, and which lies in the air gap against the stator bore. The surface A can next be reduced to the cylindrical surface of length l inside the stator bore, since everywhere else in the surface A the integrand of (7) is negligibly small. Expres-

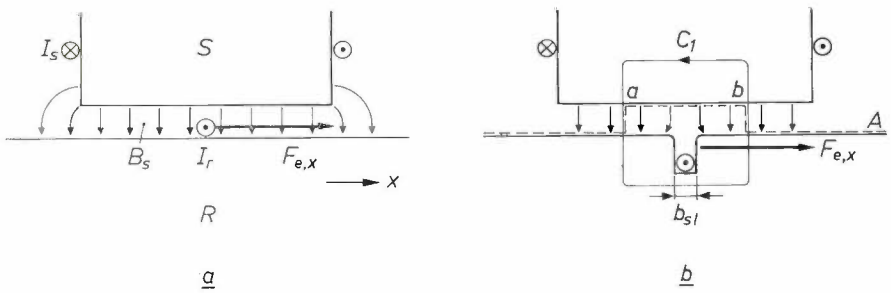


Fig. 2. A stator pole S above the rotor surface R, developed into a plane. *a*) A conductor on the rotor surface is situated in a uniform magnetic field B_s in the air gap. When a current I_r flows the conductor is subjected to a force per metre length of $F_{e,x} = B_s I_r$ in the x-direction. *b*) The rotor conductor is contained in a slot. The force per metre is again $B_s I_r$, but now acts mainly on the rotor iron. A surface and C_1 integration contour used for calculating $F_{e,x}$.

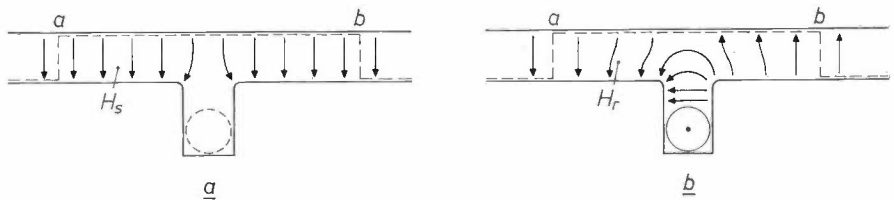


Fig. 3. *a*) The stator field H_s at a rotor slot. At points a and b the field is identical. *b*) The field H_r of a current conductor in a rotor slot. At the points a and b the field is opposite in direction.

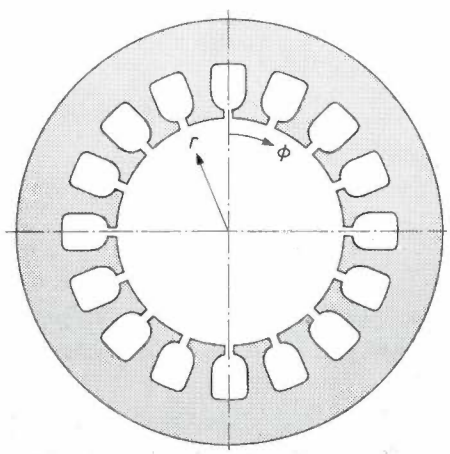


Fig. 4. Cross-section of a stator with slots. In calculating the electromagnetic torque, cylindrical coordinates r, ϕ and z are used (the z-axis is the rotation axis of the motor).

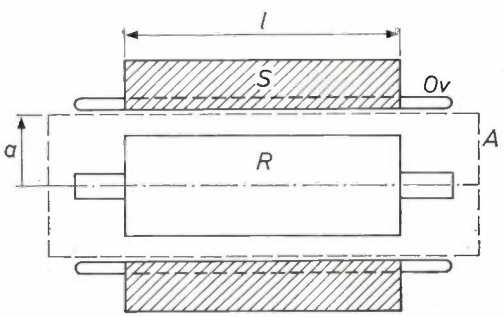


Fig. 5. In calculating the torque the Maxwell stress is integrated over an integration surface A which is located in the air gap and entirely encloses the rotor R. S stator. Ov overhang. The stator bore has a radius a and a length l in the z-direction.

sion (7) can now be simplified to

$$T_e = a^2 l \int_0^{2\pi} \mu_0 H_n H_\phi \, d\phi, \qquad (8)$$

where it is assumed that H_n and H_ϕ do not depend on z along the stator bore.

If we now wish to determine the torque with this equation, we must know the magnitude of H_n and H_ϕ at all places on the cylindrical surface to which we have reduced A. In the following we shall show that H_ϕ can be calculated from the stator currents that are accessible to measurement; H_n is a quantity that can be determined by direct measurement. These measurements are made with the field probes or measuring windings referred to earlier.

The field H_ϕ will only differ from zero at the opening of a stator slot; it contains a component $H_{s,\phi}$ due to the currents in the slot and a component $H_{r,\phi}$ due to rotor currents and also perhaps to permanent-magnet material in the rotor. The reluctance torques to which the rotor is subjected because the stator is slotted are small in the type of motor under consideration and are negligible compared with the other torque components. This implies that we may disregard $H_{r,\phi}$ and assume the remaining component $H_{s,\phi}$ to be independent of the position of the rotor.

If the stator slots are now provided with a three-phase winding and we call the phase currents $i_{s(1)}$, $i_{s(2)}$ and $i_{s(3)}$, then we can write:

$$H_\phi = H_{s,\phi} = i_{s(1)} z_{s(1)}(\phi) + i_{s(2)} z_{s(2)}(\phi) + i_{s(3)} z_{s(3)}(\phi), \qquad (9)$$

where $z(\phi)$ is the copper-density distribution (the number of wires per metre). If a slot (with opening b_{sl}) contains $N_{s(1)}$ wires that all conduct a current $i_{s(1)}$ in the same direction (see *fig. 6*), then assuming that H_ϕ is constant at the position of the slot opening, $z_{s(1)}(\phi)$ in this opening will amount to $-N_{s(1)}/b_{sl}$ wires per metre. The minus sign is a consequence of the fact that a current in the direction of the positive z-axis is regarded as positive, although the value of $H_{s,\phi}$ corresponding to it is negative. The copper-distribution functions $z_{s(1)}(\phi)$, $z_{s(2)}(\phi)$ and $z_{s(3)}(\phi)$ are thus completely determined by the magnitude and the number of the slot openings and by the way in which the turns of the phase windings are distributed among the slots and are connected with one another.

In a three-phase induction motor with $2p$ poles the functions $z_{s(1)}$, $z_{s(2)}$ and $z_{s(3)}$ are periodic in $2\pi/p$. The higher harmonic components of these functions in this type of motor give rise to unwanted (parasitic) torques. These torques, and the reluctance torques mentioned earlier, are countered as well as possible by means of a special distribution of the stator windings among the slots, by ensuring the correct ratio between the number of stator and rotor slots, and by skewing the axis of the rotor slots in relation to the stator slots.

Because of this axial skew the field H_n depends on the coordinate z. It can be shown, however, that in our case this effect can be taken into account to a very good approximation by inserting for H_n in equation (8) the values found halfway along the stator length.

In our case we shall for simplicity proceed from the initial assumption that only the fundamental harmonic components of the copper-distribution functions are significant. Assuming further that the 'magnetic axis' of phase 1 coincides with the point $\phi = 0$ (*fig. 7*), we can write:

$$\left.\begin{aligned} z_{s(1)}(\phi) &= -\hat{z}_s \sin p\phi, \\ z_{s(2)}(\phi) &= -\hat{z}_s \sin (p\phi - 2\pi/3), \\ z_{s(3)}(\phi) &= -\hat{z}_s \sin (p\phi - 4\pi/3). \end{aligned}\right\} \qquad (10)$$

Introducing these quantities into expression (9) for H_ϕ and then inserting the H_ϕ thus obtained into expression (8) for the torque, we find, after substituting B_n for $\mu_0 H_n$:

$$\begin{aligned} T_e = -a^2 l \hat{z}_s \Bigg[& i_{s(1)} \int_0^{2\pi} (\sin p\phi) B_n d\phi + \\ & + i_{s(2)} \int_0^{2\pi} \{\sin (p\phi - 2\pi/3)\} B_n d\phi + \\ & + i_{s(3)} \int_0^{2\pi} \{\sin (p\phi - 4\pi/3)\} B_n d\phi \Bigg]. \end{aligned} \qquad (11)$$

The integrals from the above relation in fact represent the value of π times the fundamental harmonic component of B_n at the successive locations $\phi = \pi/2p$, $\phi = 7\pi/6p$ and $\phi = 11\pi/6p$. A measurement of T_e that makes use of (11) therefore requires an experimental determination of these three components of B_n.

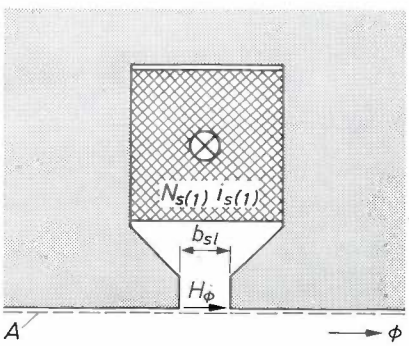


Fig. 6. Stator slot of width b_{sl}, which contains $N_{s(1)}$ wires all of which carry a current $i_{s(1)}$. H_ϕ is the tangential component of the magnetic field at the surface A located at the air gap. The copper-distribution function $z_s(\phi)$ at the location of the slot opening is $-N_{s(1)}/b_{sl}$ wires per metre; outside the slot opening it is zero.

Torque measurements by means of field probes

The three integrals in (11) can be determined experimentally by means of field probes placed in the air gap. For this purpose we used very thin Hall generators fixed to a number of the 'teeth' between the stator slots. The integrals are approximated by a summation at a limited number of points on the stator bore. For each pole pitch π/p we take N measuring points on the stator iron with each pair spaced by π/pN radians.

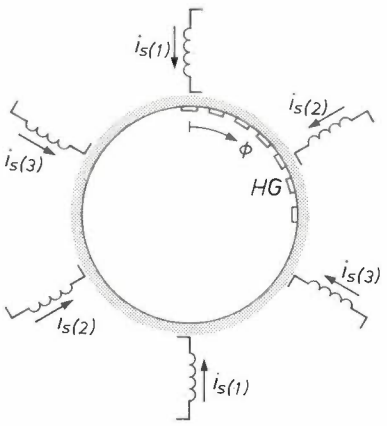


Fig. 7. Schematic cross-section of the stator of a three-phase four-pole induction motor. The coordinate ϕ is measured from the 'magnetic axis' of phase 1. One quarter of the stator circumference carries Hall generators *HG*, which are used for measuring the radial component of the field in the air gap.

With the normal assumption that B_n contains only odd higher-harmonic components, it is sufficient to locate the measuring points along the length of one pole pitch, so that we only need N measuring points. Expression (11) is then approximated by:

$$T_e \approx \frac{-2\pi a^2 l \hat{z}_s}{N}\left[i_{s(1)}\sum_{\nu=1}^{N}(\sin p\phi_\nu)\,B_n(\phi_\nu) + \right.$$
$$+ i_{s(2)}\sum_{\nu=1}^{N}\{\sin(p\phi_\nu - 2\pi/3)\}\,B_n(\phi_\nu) +$$
$$\left. + i_{s(3)}\sum_{\nu=1}^{N}\{\sin(p\phi_\nu - 4\pi/3)\}\,B_n(\phi_\nu)\right]. \qquad (12)$$

If the measurements are made on a motor with star-connected stator windings, and if the star point is not connected to a neutral, then $i_{s(1)} + i_{s(2)} + i_{s(3)}$ is always equal to zero. In that case the torque measurement can be simplified by eliminating one of the currents, for example $i_{s(1)}$. This elimination leads to the expression

$$T_e \approx -2\sqrt{3}\frac{\pi a^2 l \hat{z}_s}{N}\left[i_{s(2)}\sum_{\nu=1}^{N}\{\cos(p\phi_\nu - \pi/3)\}\,B_n(\phi_\nu) + \right.$$
$$\left. + i_{s(3)}\sum_{\nu=1}^{N}\{\cos(p\phi_\nu - 2\pi/3)\}\,B_n(\phi_\nu)\right]. \qquad (13)$$

In the theoretical case where $B_n(\phi)$ contains no higher-harmonic components the torque T_e can in principle be determined exactly from this relation, using two Hall generators ($N = 2$). In practical cases, however, there are always a large number of higher-harmonic components present because the windings are not ideal and because the stator and rotor are toothed. For an accurate measurement it is therefore desirable to take a larger value of N.

The measured results reported below were obtained on a wound-rotor induction motor. This motor has four poles ($p = 2$), 48 stator slots and 36 rotor slots. Since there are 12 stator slots for every pole pitch, the maximum accuracy would be obtained by using 12 Hall generators ($N = 12$). For practical reasons we made do with $N = 6$.

If, in the presence of higher-harmonic field components, N is nevertheless chosen equal to 2, each of these fields will in principle cause an error in the measurement. The choice of a larger N eliminates the errors attributable to certain higher harmonics: the greater the value of N, the greater the number of higher harmonics eliminated as a source of errors. The choice $N = 6$ results in the elimination of errors due to the higher harmonics of orders 3, 5, 7, 9; 15, 17, 19, 21; 27 etc. These will certainly include the by no means negligible rotor-slot harmonics of order 17 and 19.

Putting $N = 6$ in (13) gives:

$$T_e \approx \frac{-\pi a^2 l \hat{z}_s}{3}\left[i_{s(2)}\{\tfrac{1}{2}\sqrt{3}\,B_n(\pi/12) + B_n(\pi/6) + \right.$$
$$+ \tfrac{1}{2}\sqrt{3}\,B_n(\pi/4) + \tfrac{1}{2}B_n(\pi/3) - \tfrac{1}{2}B_n(\pi/2)\} +$$
$$+ i_{s(3)}\{\tfrac{1}{2}B_n(\pi/6) + \tfrac{1}{2}\sqrt{3}\,B_n(\pi/4) + B_n(\pi/3) +$$
$$\left. + \tfrac{1}{2}\sqrt{3}\,B_n(5\pi/12) + \tfrac{1}{2}B_n(\pi/2)\}\right]. \qquad (14)$$

The value of the expression between square brackets can be found directly with a measuring circuit. A slight complication is that $B_n(\pi/2)$ occurs both with a positive and with a minus sign; this makes it necessary to use an additional operational amplifier. We avoided this by also locating a Hall generator to the position $\phi = 0$ (fig. 7), since, by virtue of the symmetries in the field, $B_n(0) = -B_n(\pi/2)$. By means of three operational amplifiers (summing amplifiers) and two multipliers one can then derive from the currents $i_{s(2)}$ and $i_{s(3)}$ and the seven Hall voltages a voltage V_{hg} that is a measure of the torque T_e. *Fig. 8* shows the schematic diagram of the circuit used for the measurement. The multipliers are also Hall generators. They offer the advantage that the stator currents $i_{s(2)}$ and $i_{s(3)}$ can be made to flow directly through the current coils of the multipliers and are thus electrically isolated from the rest of the measuring circuit.

Results of the measurements

Fig. 9 shows a torque-speed characteristic, in which the signal V_{hg} is plotted as a function of the speed n (n_0 is the synchronous speed). Fig. 9*a* gives the unfiltered signal V_{hg}, and fig. 9*b* the signal V_{hg} when a lowpass filter with a cut-off frequency of 20 Hz is used. Fig. 9*a* shows that, although seven Hall generators were used, the measured signal still contains a large number of unwanted signals owing to the unsuppressed higher-harmonic fields. It can be seen from fig. 9*b* that the filter is not effective at very low speeds. The explanation for this is that the unwanted signals then contain low-frequency components due to the rotor slots rotating at low speed.

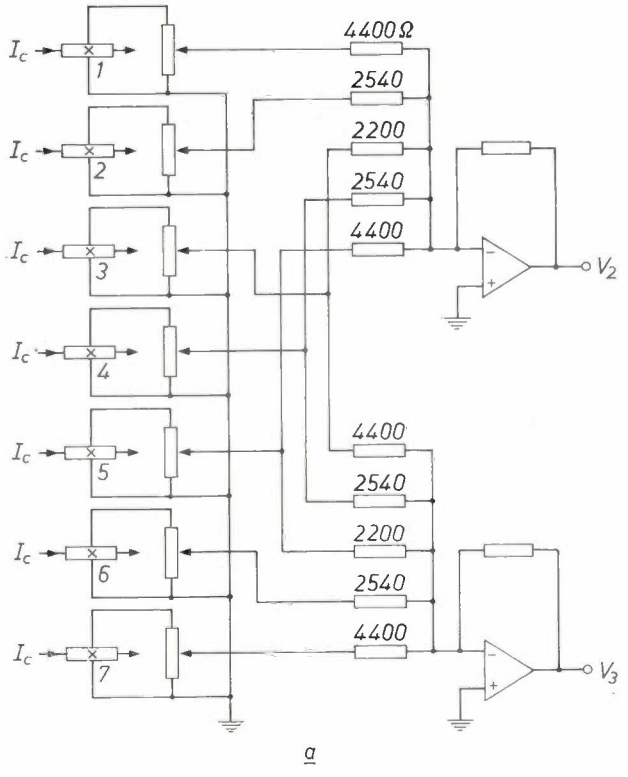


Fig. 8. Circuit for measurements with Hall generators. *a*) Adding circuit in which the expressions between curly brackets in equation (14) are summed. A current I_c flows through each of the Hall generators *1-7*. The output signals, after weighting in accordance with the coefficients of equation (14), are added by means of two operational amplifiers to form the voltages V_2 and V_3. *b*) The voltages V_2 and V_3 are multiplied by the value of the stator currents $i_{s(2)}$ and $i_{s(3)}$ in two further Hall generators; the currents in two coils produce the magnetic field in which the Hall generators are located. After further addition an output voltage V_{hg} results, which is proportional to the torque of the motor.

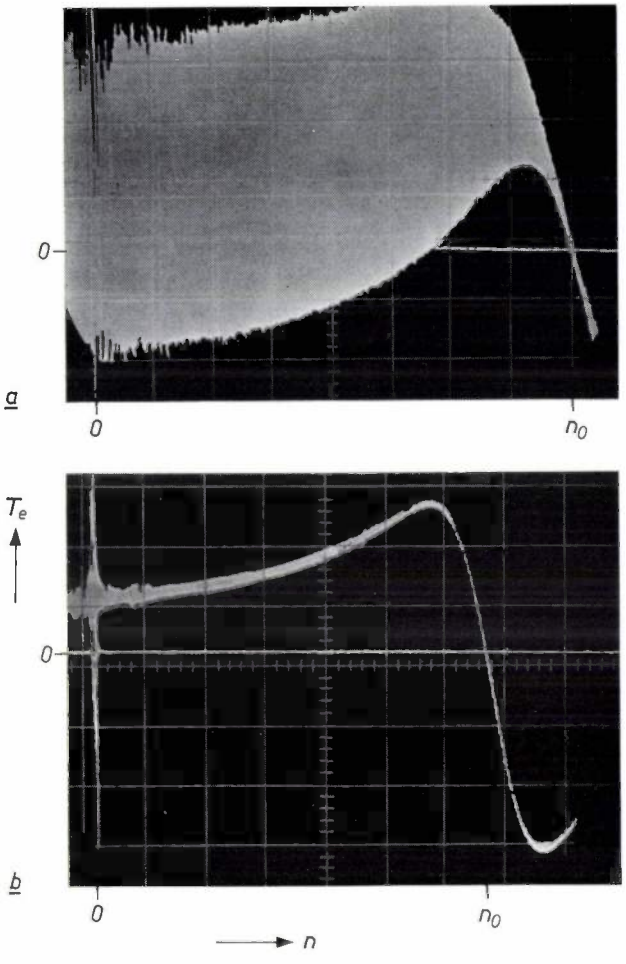


Fig. 9. Torque-speed curve measured with Hall generators, *a*) unfiltered and *b*) filtered. T_e electromagnetic torque of the motor. n speed, n_0 synchronous speed. The lowpass filter used in (*b*) has a cut-off frequency of 20 Hz.

Fig. 10a shows the signal V_{hg} as a function of time during no-load starting. In fig. 10*b* the same effect is recorded, but now making use of a lowpass filter with a cut-off frequency of 500 Hz. The torque pulses occurring during this transient effect have a frequency of about 50 Hz. These 50-Hz pulses are caused by the fact that when the motor is switched on d.c. transient components (equalizing currents) are set up in the windings of the motor. The interaction of these d.c. components with the 50-Hz rotating field causes the noticeable 50-Hz torque pulses.

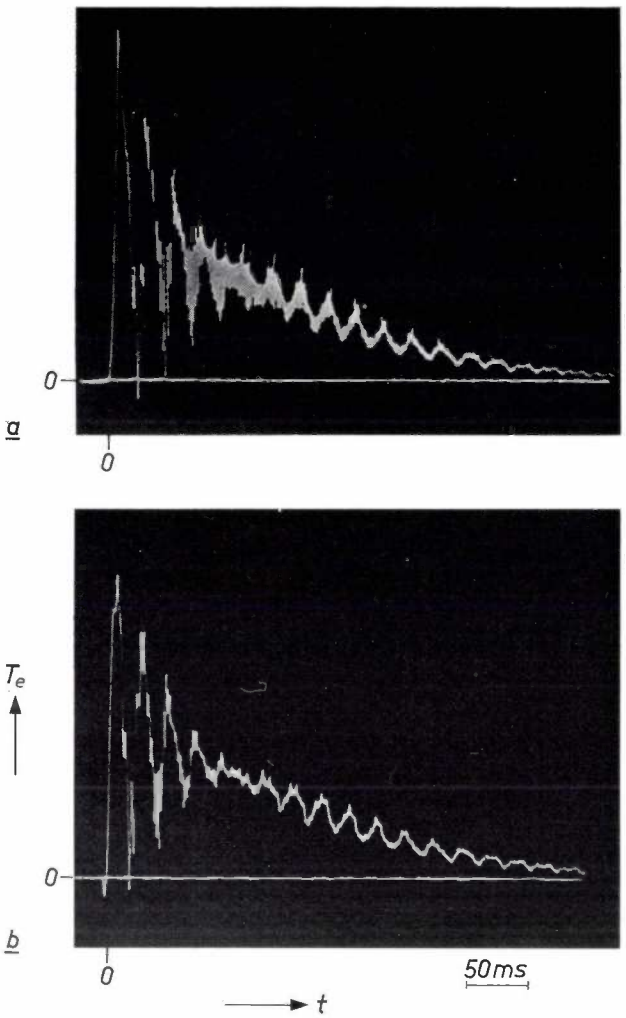


Fig. 10. Torque curve measured with Hall generators during no-load starting, *a*) unfiltered and *b*) filtered. The 50 Hz pulses on the torque are attributable to the d.c. surges that flow in the motor windings when the motor is switched on directly. The lowpass filter used in (*b*) has a cut-off frequency of 500 Hz.

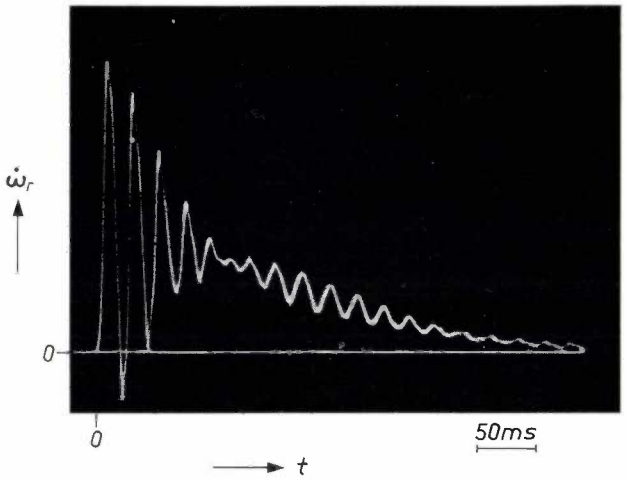


Fig. 11. Angular acceleration $\dot{\omega}_r$ of the rotor measured with an angular-acceleration meter during a no-load start. The curve shows good agreement with fig. 10.

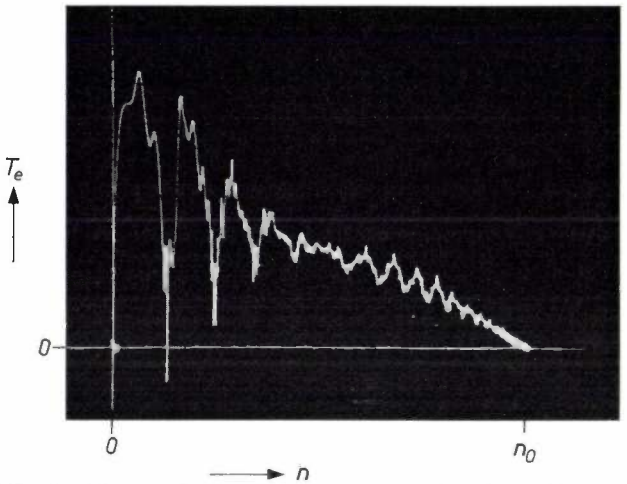


Fig. 12. Measured curve of the torque as a function of motor speed during a no-load start (dynamic torque-speed curve). The great difference as compared with the static torque-speed curve (fig. 9) is attributable to the d.c. currents.

In the case of an unloaded motor with a very small frictional torque T_{fr} the angular acceleration $\dot{\omega}_r$ is approximately proportional to the torque T_e. To verify our measurement method we measured the angular acceleration $\dot{\omega}_r$ with an angular acceleration meter [2] mounted on the motor shaft. The result is shown in *fig. 11*. Comparison of this with fig. 10 shows that there is good agreement.

Finally, during a no-load start, the signal V_{hg} was recorded as a function of the angular velocity. The resultant dynamic torque-speed curve is shown in *fig. 12*. The curve is quite different in shape from the static torque-speed curve; this is also related to the transient equalization currents that occur when the motor is switched on and certainly last as long as the no-load start represented here. For comparison, *fig. 13* shows the same effect in a plot of the angular acceleration as a function of angular velocity. Here again the agreement is good.

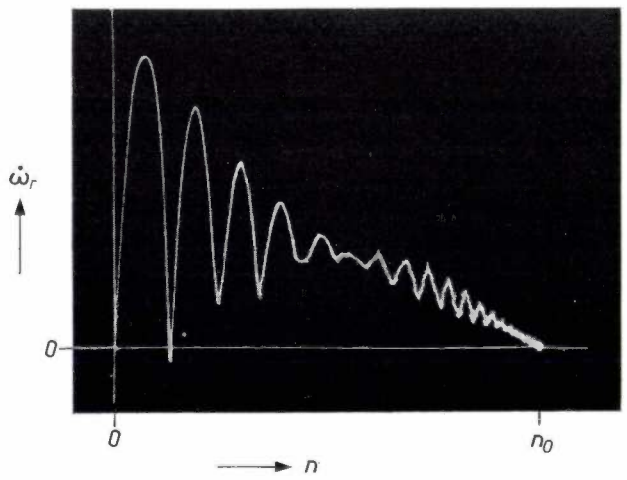


Fig. 13. Angular acceleration measured during the experiment in fig. 12. The natural frequency of the angular acceleration meter is too low to reproduce all fast torque variations.

[2] Made by Dr Staiger, Mohilo and Co. (type 1326).

Torque measurements using measuring windings

For measuring the torque with windings instead of with Hall generators we proceed from the same equation (11), but write it in the form

$$T_e = i_{s(1)}\Phi_{(1)} + i_{s(2)}\Phi_{(2)} + i_{s(3)}\Phi_{(3)}, \qquad (15)$$

where

$$\Phi_{(1)} = -a^2 l\hat{z}_s \int_0^{2\pi} (\sin p\phi) B_n \mathrm{d}\phi,$$

$$\Phi_{(2)} = -a^2 l\hat{z}_s \int_0^{2\pi} \{\sin (p\phi - 2\pi/3)\} B_n \mathrm{d}\phi,$$

and

$$\Phi_{(3)} = -a^2 l\hat{z}_s \int_0^{2\pi} \{\sin (p\phi - 4\pi/3)\} B_n \mathrm{d}\phi.$$

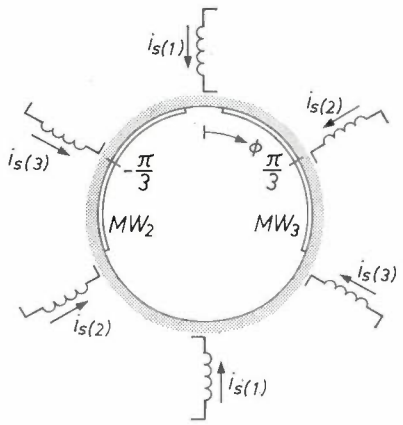


Fig. 14. The measuring windings MW_2 and MW_3 in the stator of a three-phase four-pole induction motor. The windings are located symmetrically with respect to the points $\phi = -\pi/3$ and $\phi = \pi/3$.

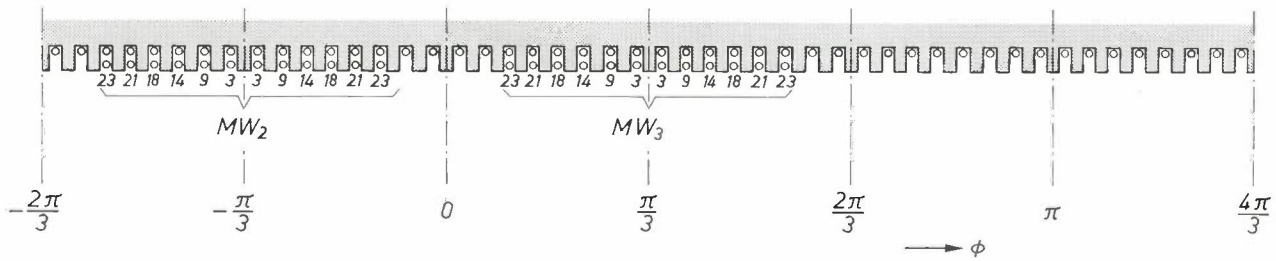


Fig. 15. View of the stator winding with the measuring windings MW_2 and MW_3. The numbers of wires are indicated.

The quantities Φ have the dimension of a magnetic flux.

Now a magnetic flux linked by a winding can be measured by an integration over time of the voltage e induced in the winding, since Faraday's second law of induction states that $e = -\mathrm{d}\Phi/\mathrm{d}t$, and therefore

$$\Phi(t) = -\int_0^t e\mathrm{d}t' + \Phi_0,$$

where Φ_0 is the value of Φ at the time $t = 0$.

It is easily verified that a measuring winding with a copper-distribution function $z_{m(1)} = p\hat{z}_m \cos p\phi = -p\hat{z}_m \sin (p\phi - \pi/2)$ links a flux which is proportional to $\Phi_{(1)}$, while copper distributions $z_{m(2)} = -p\hat{z}_m \sin (p\phi - 7\pi/6)$ and $z_{m(3)} = -p\hat{z}_m \sin (p\phi - 11\pi/6)$ will link a flux proportional to $\Phi_{(2)}$ and $\Phi_{(3)}$, respectively. It is then assumed, however, that the flux due to the current in the measuring winding and the stray flux linked by the overhang (i.e. the end-connections) of the measuring winding are negligibly small. This can be accomplished in practice by keeping the load on the measuring windings very low and the overhang as small as possible.

Elimination of the current $i_{s(1)}$ from equation (15) gives the following simplification:

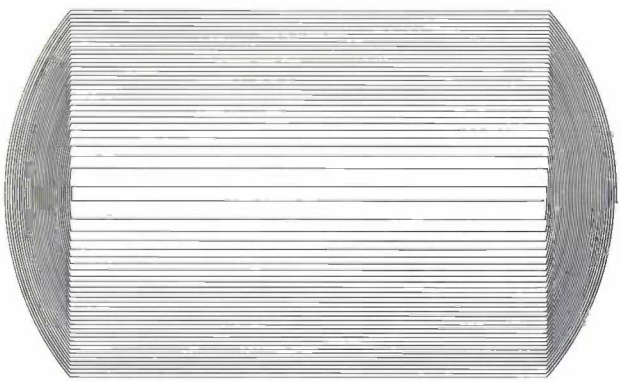


Fig. 16. Measuring winding in the form of printed wiring on plastic film.

$$T_e = i_{s(2)}\Phi_{(21)} + i_{s(3)}\Phi_{(31)}. \qquad (16)$$

Fluxes proportional to $\Phi_{(21)} = \Phi_{(2)} - \Phi_{(1)}$ and $\Phi_{(31)} = \Phi_{(3)} - \Phi_{(1)}$ are linked by measuring windings with copper-distribution functions

$$z_{m(21)} = \sqrt{3}p\hat{z}_m \sin (p\phi + 2\pi/3) \text{ and}$$
$$z_{m(31)} = -\sqrt{3}p\hat{z}_m \sin (p\phi - 2\pi/3).$$

By virtue of the existing symmetries, each measuring winding only needs to extend over one pole pitch ($= \pi/p$ radians). Since the motor is a four-pole type ($p = 2$), we can link a flux that is proportional to

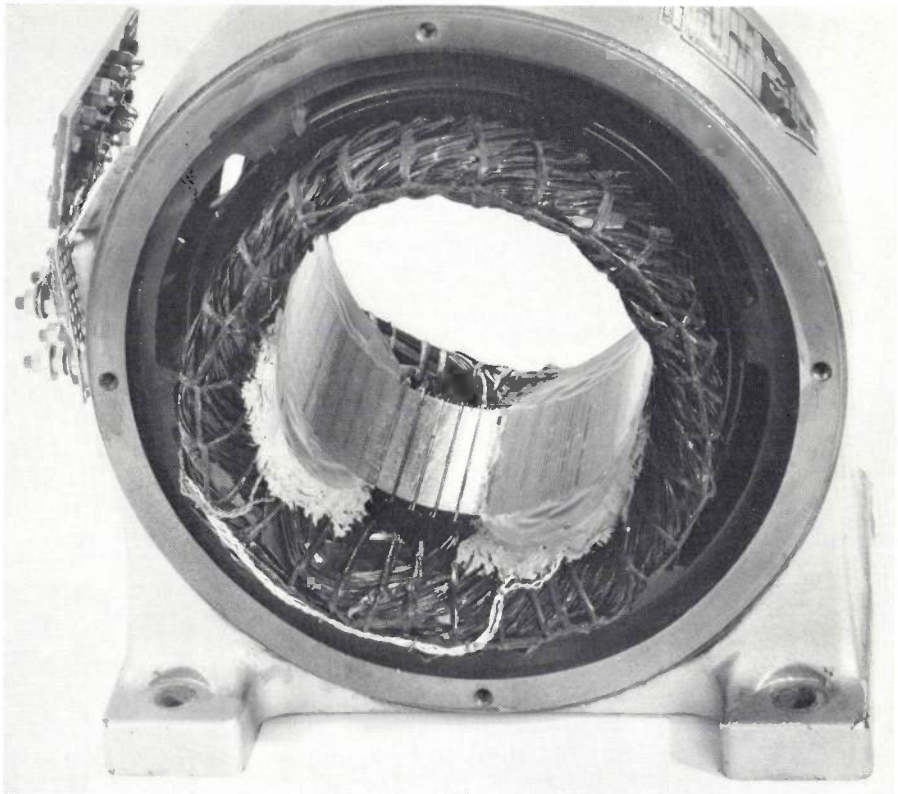


Fig. 17. Stator of an 11 kW (15 hp) three-phase induction motor. Two measuring windings on film are fitted in the air gap.

$\Phi_{(21)}$ with a measuring winding MW_2 that extends over $\pi/2$ radians and is located symmetrically with respect to the point $\phi = -\pi/3$, and we can link a flux that is proportional to $\Phi_{(31)}$ with a measuring winding MW_3 that is located symmetrically with respect to the point $\phi = \pi/3$ (*fig. 14*).

The measuring windings are made of thin copper wire. In order to distribute the windings as sinusoidally as possible the sides of the coils are arranged as shown in *fig. 15* in the openings of the existing stator slots. Another possibility is to apply the measuring winding in the form of printed wiring to a plastic film (*fig. 16*) and to introduce this into the air gap (*fig. 17*).

The measuring circuit is illustrated schematically in *fig. 18*. It contains two operational amplifiers arranged as integrators, a summing amplifier and two Hall generators for multiplication by the stator-current values. The symbol V_{mw} denotes the measured signal.

Results of the measurements

To facilitate comparison, the same measurements that were carried out with field probes were also made with measuring windings. *Fig. 19* — to be compared with fig. 9 — shows that the unwanted signals remain within reasonable bounds even in the unfiltered signal V_{mw}.

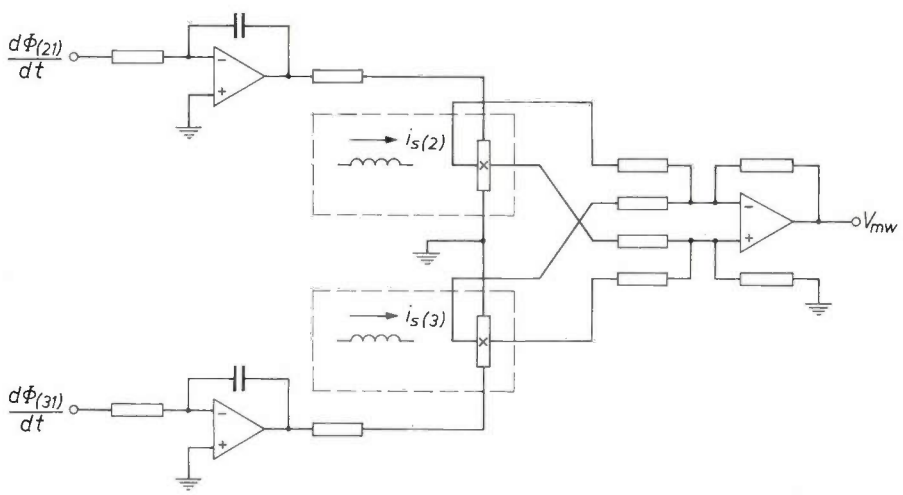


Fig. 18. The circuit used with the measuring windings. The voltages $d\Phi_{(21)}/dt$ and $d\Phi_{(31)}/dt$ from the measuring windings are applied to two integrators (comprising an operational amplifier) and are then multiplied by the stator currents $i_{s(2)}$ and $i_{s(3)}$ with the aid of Hall generators. The results are added to give the output signal V_{mw}, which is proportional to the torque of the motor.

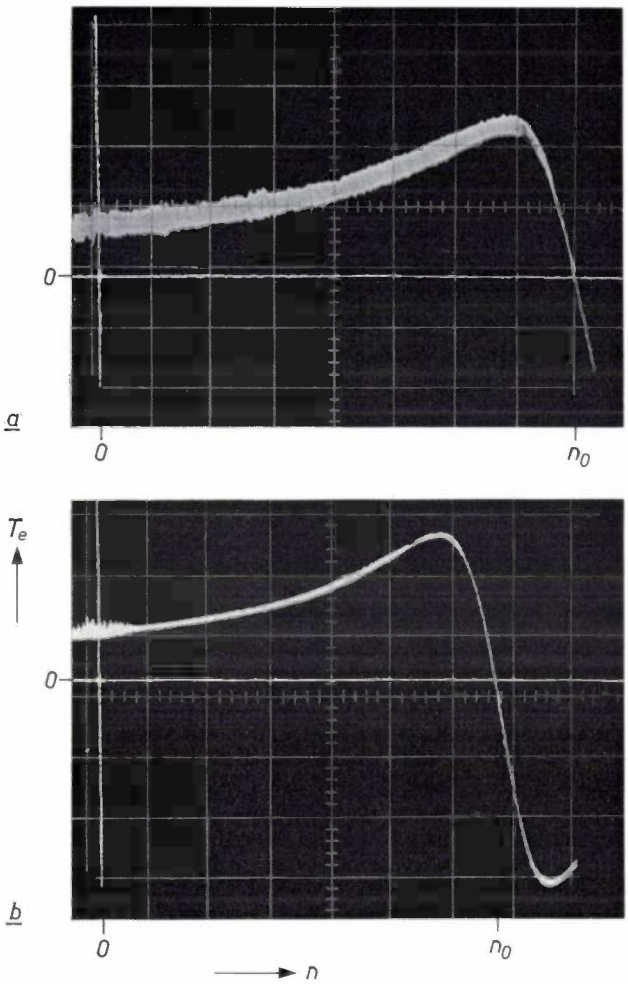


Fig. 19. Torque-speed characteristic recorded with measuring windings, *a*) unfiltered and *b*) filtered. T_e electromagnetic torque of the motor. The lowpass filter used in (*b*) has a cut-off frequency of 20 Hz. Compared with the measurement using Hall generators (fig. 9) there is less interference from unwanted signals.

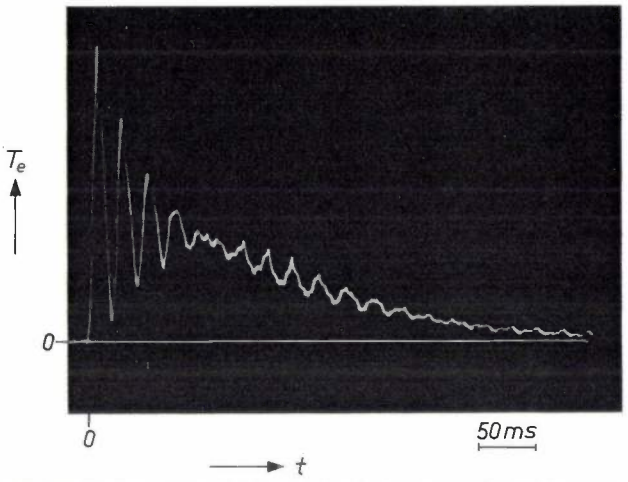


Fig. 20. Torque curve during no-load start, recorded with measuring windings. Although no lowpass filter was used, there is less interference than in the corresponding measurement using Hall generators (fig. 10).

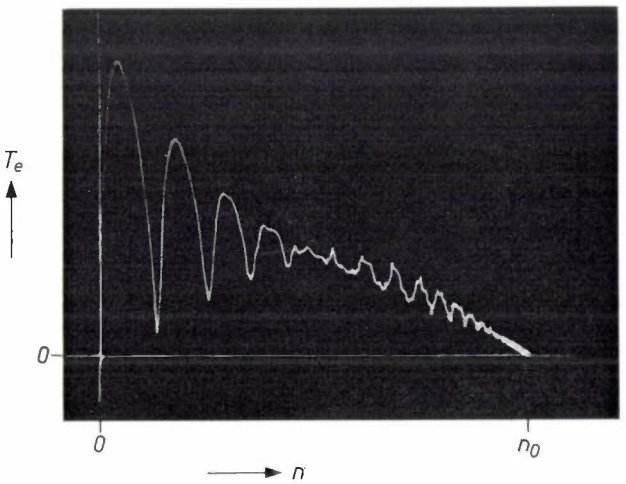


Fig. 21. Dynamic torque-speed characteristic recorded with measuring windings. Again, there is less interference than in the trace made with Hall generators (fig. 12).

The results obtained with measuring windings do not show the perturbations visible in the curves measured with Hall generators because the distribution over the stator circumference is much more uniform; the interference from higher harmonics due to the fairly coarse sampling of the magnetic field with just a few Hall generators is not present in this case. *Figs. 20* and *21*, recorded without using a filter, show a close resemblance to figures 10 and 12.

It can be seen that torque measurements using windings yield better results than those using field probes. Added to this is the fact that field-probe measurements can be difficult because of the small radial dimension of the air gap. For these reasons windings are preferable to probes for torque measurements on a.c. motors.

Summary. The torque generated at any given instant by an electric motor depends on the magnetic field in the air gap at that instant. The tangential component of this field can be derived from the currents in the windings; the radial component can only be determined with sufficient accuracy by measurements. These measurements can be carried out by means of Hall generators or windings introduced into the air gap, which also make it possible to record the torque during fast transients. Trials with both types of measurement on a three-phase induction motor have shown that the measuring windings give the best results.

A speed-controlled d.c. motor for a washing machine

R. Raes and J. Schellekens

The washing machine

Nearly all modern automatic washing machines have one drum, which rotates about a horizontal axis and is used both for washing and spin-drying. This means that the drum must be capable of rotating at widely different speeds. If we look only at the speed, a typical wash programme runs as follows:

1. Wash or rinse; drum speed 50 revolutions per minute. The load (i.e. the washing) is carried round by the drum and pulled out of the water partly filling the drum; it then drops back from the uppermost point of the drum.
2. Drain; nominal speed 75 to 90 rev/min. The water is slowly pumped away, so that the braking torque decreases and the number of revolutions slowly increases from the wash speed to the drain speed. The load gradually collects around the circumference of the drum and distributes itself uniformly: the pieces nearest the centre remain longest in movement and eventually settle on the unoccupied parts of the circumference.
3. Spin; speed in the range from 400 to 750 rev/min. In some cases the nominal speed is adjustable, so that it can be adapted to the type of load and the amount of drying required.

In view of the considerable difference in speed a d.c. motor is evidently the type of motor required in a washing machine, since its speed can so easily be controlled. In this article we shall describe a d.c. motor that can provide the required drum speed by means of a fixed transmission, together with an electronic circuit that automatically controls the speed of the motor. This system has been developed by the MBLE motors group for a washing machine that spin-dries at 500 rev/min. The system can easily be modified to give a spin-drying speed of 750 rev/min.

Philips have been marketing washing machines fitted with this motor for more than a year. The electronic circuit in these machines differs somewhat from that described here, but is based on the same principle; it was developed by the Philips factories in Amiens and Evreux.

Ir R. Raes and J. Schellekens, Techn. Ing., are with S.A. Manufacture Belge de Lampes et de Matériel Electronique (MBLE), Brussels.

Motor and power supply

In all stages of the development of the motor we were guided by the need for it to be rugged and easy to manufacture. The motor had to have a life of 4000 hours, including 400 hours at the highest speed; this corresponds to about 2500 complete wash programmes. We decided to use a motor with permanent-magnet excitation, and thus avoided the necessity for an expensive winding process for the stator.

Although in practice the motor is supplied with pulses of current, we shall base our discussion in the first place on the simple equations applicable for d.c. supply [1]:

$$E = k_1 N \Phi n, \tag{1}$$

$$T_e = k_2 N \Phi I, \tag{2}$$

$$V = E + RI, \tag{3}$$

$$P = EI. \tag{4}$$

Here E is the speed voltage (or back e.m.f.), N the number of rotor windings, Φ the stator flux through the rotor windings, n the motor speed, I the rotor current, T_e the torque acting on the rotor, V the voltage across the motor, R the internal resistance of the rotor circuit and P the delivered power; k_1 and k_2 are proportionality constants. (If n is expressed in revolutions per minute, then $k_1/k_2 = 2\pi/60$.)

The motor is supplied from the mains through a full-wave bridge rectifier, consisting of two thyristors and two diodes (*fig. 1*). The electronic circuit *El* determines the triggering point t_0 of the thyristors in each half-cycle of the mains voltage, thereby determining the power supplied and hence the motor speed. Before triggering, i.e. at the beginning of each half-cycle, the current I is zero; equations (1) and (3) show that the voltage V across the motor is then a direct measure of the speed n. The speed can therefore be stabilized by feedback of V to the electronic circuit.

The series resistance R_s in the circuit shown in fig. 1 combines a number of functions [2]. In the first place, R_s is used for heating the washing water; this is done

[1] See for example the article by E. M. H. Kamerbeek in the first issue devoted to small electric motors, Philips tech. Rev. 33, 215, 1973 (No. 8/9), in particular page 231.

[2] W. Ebbinge and D. C. de Ruiter, Electronic Appl. 29, 29, 1969.

with the switch S_1 closed, the motor then being switched off. In the second place, when S_1 is open, R_s limits the peak value of the current in the semiconductors. Thirdly, when R_s is combined with the circuit in fig. 1 — which has no transformers and only simple electronics — power supply to the motor and its speed control are both satisfactory under the widely different conditions of washing, draining and spin-drying. The heat generated in R_s during the washing process is taken up by the washing water. During spin-drying there is no water in the machine, but then much less heat is developed by R_s.

The following very rough calculation will make this clearer. The power required for washing (at 50 rev/min) is about 75 W, while that required at the highest spin-drying speed (750 rev/min) is about 200 W. During spin-drying (high speed, low torque) the speed voltage is high and the current low. The speed voltage E is about 190 V. Most of the mains voltage is therefore used for compensating the speed voltage; R_sI and RI are small. During washing the speed and therefore the speed voltage are 15 times lower: $E \approx 12.5$ V. To obtain the required power of 75 W the current through the motor must therefore average 6 A. It follows that the internal resistance must be small, otherwise the heat generated in the motor would be excessive; in our motor $R = 0.9\ \Omega$. The voltage across the motor during washing should then average about 18 V (i.e. $12.5 + 0.9 \times 6$ V). This means that in the absence of R_s the voltage from the mains would be much too high. Although the correct power level could in principle be obtained by triggering the thyristors very late in each half-cycle, in practice this is unsatisfactory because it gives a very poor 'form factor' for the current. Now, however, the resistance R_s, which is 23 Ω, accounts for a large part of the voltage drop, $6 \times 23 \approx 140$ V.

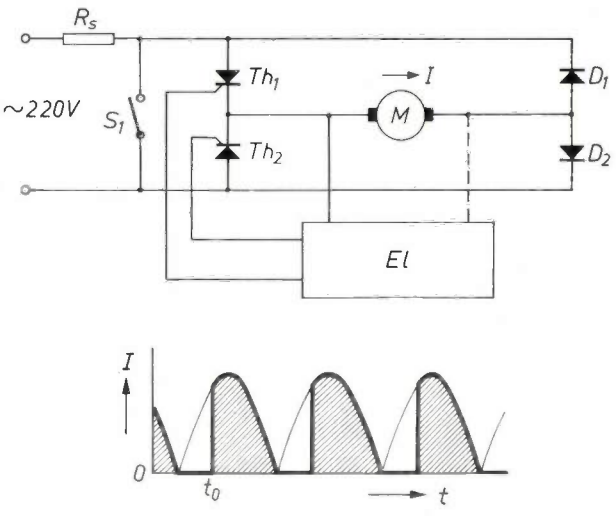


Fig. 1. Voltage supply and speed control of the motor, schematic. *Above:* power is supplied to the motor M from the mains via a bridge rectifier consisting of two thyristors, Th_1, Th_2, and two diodes, D_1, D_2. *Below:* the current through the motor as a function of time. In each mains half-cycle the motor only receives current from the instant t_0 (the 'triggering point') at which one of the thyristors is triggered by the electronic circuit *El*. The connection indicated by the dashed line provides the electronic circuit with an indication of the speed, via the speed voltage, for stabilizing the speed. The series resistor R_s limits the current in the semiconductors and also permits the thyristors to be triggered early in each mains half-cycle both for washing and spin-drying; late triggering would adversely affect the form factor of the current. The same resistor is used for heating the water when S_1 is closed.

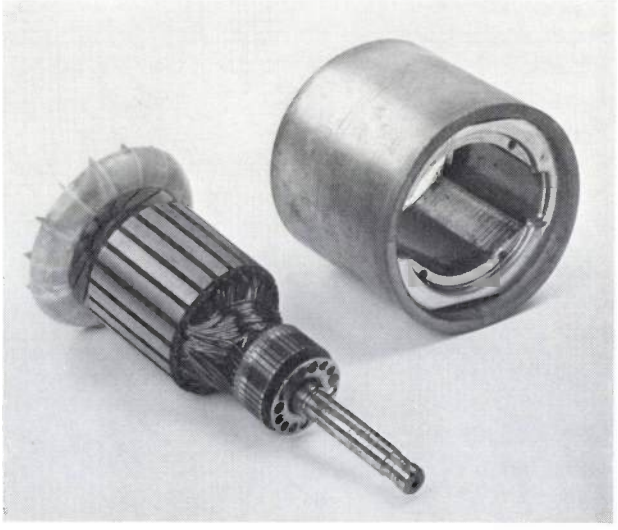


Fig. 2. The rotor and the stator.

A fast-running motor is preferable to a slow-running one because it can be smaller for a given delivered power. The maximum speed of our motor is 12 000 rev/min, which is 16 times the greatest drum speed, 750 rev/min. We have been able to obtain this large transmission ratio in one step by means of a 'poly-vee belt', a driving belt whose profile consists of a number of vees (in our case six) and which is sufficiently flexible yet gives an adequate grip on the small pulley (which has a corresponding profile). The motor speed for washing is then $50 \times 16 = 800$ rev/min. The speed voltage varies from 12 V at 800 rev/min to 190 V at 12 000 rev/min.

The critical point affecting the life of the motor is commutation. Too little attention paid to this point can easily lead to sparking and pitting of the commutator, and it is all the more important here because of the high terminal voltage at the highest running speeds and the pulsed nature of the supply. To prevent the occurrence of undesirable voltage pulses at the commutator contacts and to keep the unavoidable voltage pulses as small as possible, we tried to keep the distribution of the stator flux as smooth as possible and the inductance of the commutating coil in the rotor as low as possible.

We shall first take a closer look at the stator and the rotor (see *fig. 2*), and we shall then discuss the electronics in greater detail.

The stator

Fig. 3 shows the stator in cross-section. It consists of two magnets *M* of ferroxdure (Fxd 280 K) contained in an iron housing *H* and provided with iron pole pieces *P*. The magnetic flux passes from one magnet to the other through the pole pieces and the rotor, and then back through the housing again.

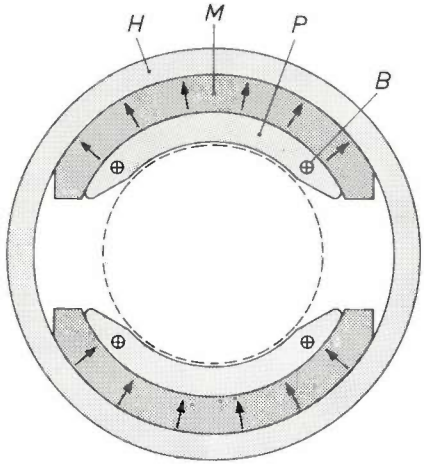


Fig. 3. Cross-section of the stator. *H* soft-iron housing. *M* magnets of ferroxdure Fxd 280 K. *P* pole pieces of soft-iron laminations positioned by iron rods *B*. The dashed line indicates the rotor. The housing has an outside diameter of 110 mm and is 91 mm long.

iron rods *B*. The thickness of the ferroxdure magnets is made large enough to prevent demagnetization in the event of overloading or brief short-circuiting of the rotor terminals when the motor is running.

The housing is an iron ring with a wall thickness of 8 mm, an outside diameter of 110 mm and a length of 91 mm, cut from tube of the appropriate dimensions.

In the assembly of the stator the magnets and pole pieces are positioned in the housing by means of aluminium flanges. The components are then cemented together, and the stator is then provided with a second set of flanges which will later carry the rotor ball-bearings and serve as attachment points for the suspension system of the motor in the washing machine.

The stator is first fully magnetized on a special yoke. Later, after assembly of the complete motor, demagnetization brings the flux through the rotor to the required value (15.5×10^{-4} Wb or 15.5×10^{4} maxwells), corresponding to a speed voltage of 16 V at 1000 rev/min.

The rotor

The rotor (*fig. 4*) has 22 slots. For each slot there are *two* coils and consequently two segments on the commutator *C*, so that there are altogether 44 segments. The large number of slots and segments minimizes the

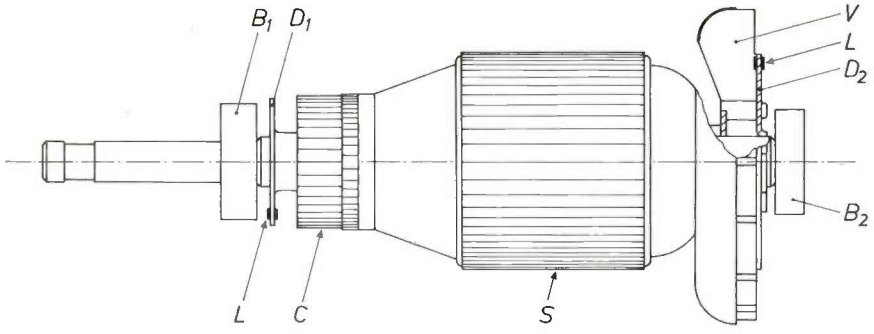


Fig. 4. The rotor. *S* rotor surface with slots. *C* commutator. *V* ventilator fan. B_1,B_2 ball-bearings. The rotor is balanced by lead slugs *L* pressed into the discs D_1 and D_2. One of these discs, D_1, also keeps grease from the ball-bearing B_1 away from the commutator, while D_2 provides the surface for the attachment of the plastic fan, keeping it out of direct contact with the hot motor shaft.

The magnets are 73 mm long, and are the largest yet made from a single piece of ferroxdure. The large surface, at a given magnetization, delivers a high flux. The pole pieces ensure that the flux is distributed effectively, i.e. roughly sinusoidally, along the circumference of the rotor. In addition a smooth inside surface contributes towards a smooth distribution of the flux. Imperfections in the ferroxdure (which is difficult to machine) are thus compensated by the pole pieces. These are laminated, with the laminations located by means of

inductance of each commutating coil. It also has the advantage that the ripple on the speed voltage is low. This is necessary for good speed stabilization, since the speed voltage is used to give an indication of the motor speed.

The rotor has two special discs D_1 and D_2 for balancing. The balance is adjusted by lead slugs *L* pressed into holes in the discs. With this method of balancing there is no need to drill holes into the rotor, which would alter its magnetic properties.

The rotor is also fitted with a plastic radial ventilator fan V. One of the balancer discs (D_2) also acts as an air guide and as the attachment for the fan, so that the plastic does not come into direct contact with the hot motor shaft. The other disc (D_1) has the secondary function of keeping the grease from the ball-bearing B_1 away from the commutator.

The rotor is wound in an automatic process, in which the ends of the wires are also automatically pressed into the commutator slots.

The power section (the motor and the bridge rectifier for the voltage supply) is enclosed in fig. 5 by the dashed line P. The section includes a suppressor C_1R_1 that prevents the steep pulses from the circuit from reaching the supply network and also prevents voltage peaks on the mains from damaging the circuit. The diodes D_1 and D_2 also form part of a bridge rectifier $D_1D_2D_3D_4$ which supplies the voltage for the control circuit. The supply current for the circuit flows from a through the circuit to b, and through the motor to c.

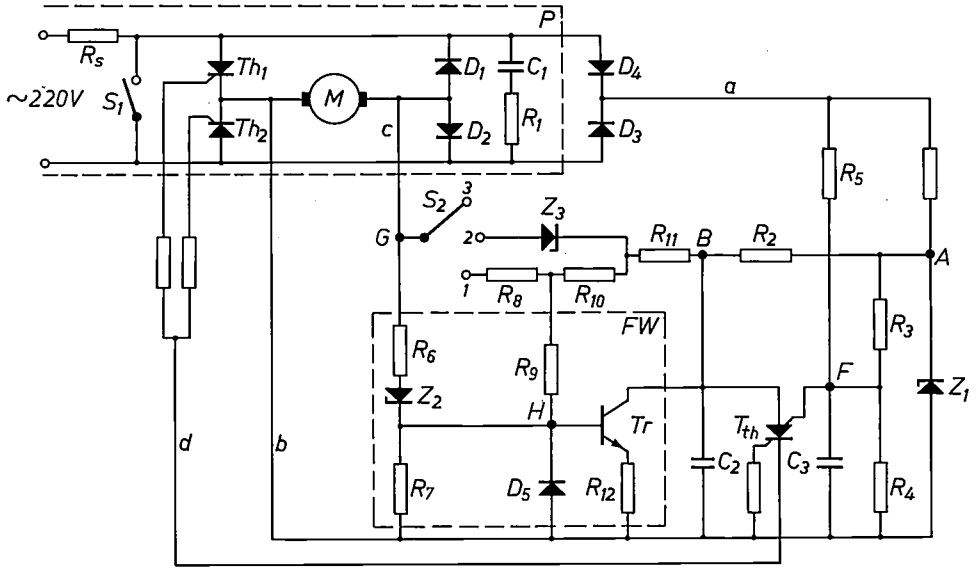


Fig. 5. The electronic circuit that controls the motor speed. The dashed line P indicates the power section. Positions of the switch S_2: *1* wash, *2* drain, *3* spin-dry. In each half-cycle of the mains, C_2 is recharged from A; when the voltage across C_2 reaches that of F, then C_2 discharges through T_{th} and one of the thyristors Th_1, Th_2 is triggered.

The electronic circuit

As mentioned earlier, the electronic circuit (*El* in fig. 1) determines the point in each half-cycle at which the thyristors are triggered and hence the speed of the motor. The schematic circuit diagram is shown in *fig. 5*, together with the motor and the voltage supply. When switch S_2 is in position *1* the motor is set for 'wash', position *2* is for 'drain' and position *3* for 'spin'. This switch is itself actuated by a programme selector, which controls the complete wash programme. In positions *1* and *2* the speed is stabilized, in position *3* it is not. A safety feature is that in the event of a power failure, or if the motor has been jammed for some reason or other, the motor cannot be restarted in positions *2* and *3* but must first go through the wash cycle again. This is necessary to prevent any part of the load that has fallen to one side from being brought up to spin-drying speed before the load has been evenly distributed again during 'drain', which would put the drum dangerously out of balance. We shall now discuss the general operation of the circuit.

The line b forms the 'earth' of the circuit; since the cathodes of the thyristors Th_1 and Th_2 are also connected to this line, they can be triggered directly from the circuit (via d), without for example having to use a pulse transformer. If required, the speed voltage can be fed back to the circuit through G.

We shall first consider what would happen, with S_2 in position *3*, if the part inside the dashed line *FW* were missing. At the end of a half-cycle of the mains voltage the thyristors are turned off, and in the next half-cycle the voltage across the circuit rises rapidly until A has reached the Zener voltage (47 V) of the Zener diode Z_1 (*fig. 6*). With S_2 in position *3* the branch on the left of B in fig. 5 is inoperative, and C_2 begins to charge up (V_B in fig. 6) with the RC time constant of C_2R_2 (about 1 millisecond). B is also one of the inputs of the 'thyristor tetrode' T_{th} (two transistors in a thyristor configuration). Point F is the other input. V_F is a fixed fraction of V_A determined by R_3R_4. The thyristor tetrode acts as a 'switch' which changes state when V_B becomes greater than the 'reference voltage' V_F. At the

instant when this happens (t_0 in fig. 6) C_2 discharges through d to the trigger electrodes of the thyristors Th_1 and Th_2, and one of them — the one with the positive anode-cathode potential — is triggered. From t_0, therefore, power is again supplied to the motor. Since the triggered thyristor forms a short-circuit between the now positive pole of the mains and the 'earth' b, the voltage across the control circuit vanishes completely at the point t_0. Thus in every half-cycle a completely fresh start is made, unaffected by the previous cycle. Consequently, when S_2 is in position *3* the triggering point t_0 is essentially determined by the voltage divider R_3R_4 and by the RC constant of R_2C_2. This time constant is made so small that t_0 falls close to the beginning of each half-cycle, so that very nearly full advantage is taken of the mains voltage.

The resistor R_5 stabilizes the circuit against mains fluctuations. If, for example, the mains voltage is a little too high, which would cause the motor to run too fast, the effect of the connection through R_5 is to make V_F slightly higher than the fixed fraction of V_A determined by R_3R_4. This retards the triggering point in the fig. 6 configuration and thus reduces the supplied power.

The capacitor C_3 acts as a filter. At the highest speed the motor has a fairly large high-frequency ripple voltage, which also appears, after some attenuation, at A. The corresponding ripple on the reference voltage V_F would cause some uncertainty in the triggering point. C_3 filters this ripple out of the reference voltage. The combination R_5C_3 also improves the shape of the curves in fig. 6, but this will not be dealt with here.

When S_2 is set to position *1*, the resistors between A and G form a voltage divider which lowers the 'target voltage' of B (the voltage which B would reach if T_{th} did not switch on): when the motor is running, V_G is lower than V_b before the triggering of a thyristor, and is therefore certainly lower than V_A (the speed voltage is $V_b - V_G$). It can be seen from fig. 6 that this leads to later triggering and therefore to a lower motor speed. In qualitative terms it is easy to see that the effect of this is to stabilize the motor speed. Before a thyristor is triggered, V_G (which is negative) is proportional to the speed. If for example the speed should now become less than the desired nominal value, V_G then becomes less negative, the target voltage of B becomes higher, the triggering is advanced and more power is supplied, so that the speed is raised again. It is evident that in position *3* the speed is not stabilized.

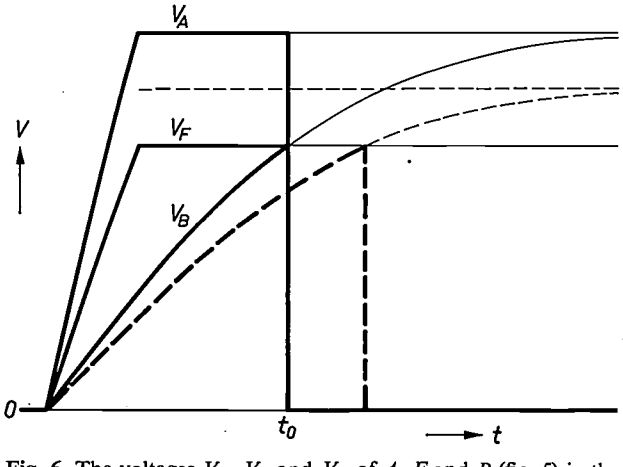


Fig. 6. The voltages V_A, V_F and V_B of A, F and B (fig. 5) in the first part of each half-cycle of the mains voltage, when S_2 is in position *3*. V_A rises quickly to 47 V, the Zener voltage of Z_1; V_F is a fixed fraction of this. V_B approaches V_A with the time constant of R_2C_2, but when V_B exceeds V_F, T_{th} switches on and one of the thyristors is triggered. From that moment (t_0) the motor receives current, and the voltages in the circuit drop back to zero. Dashed line: V_B and the 'target voltage' of B in position *1* of S_2. When the motor is running, V_G is negative and the target voltage of B is therefore lower than V_A. The triggering then occurs later.

When the motor is not running and S_2 is in position *1* the situation is not quite so straightforward. Suppose that the mains voltage has returned to zero after a half-cycle in which Th_1 has been conducting. Because of the inductance of the motor, a current will still flow briefly in the circuit MD_1Th_1 that will prevent Th_1 from switching off. As a result, c acquires a *positive* potential with respect to b, equal to the sum of the forward voltages of D_1 and Th_1 (about 2 V). This is referred to as the *flywheel effect*. Although the effect is very short-lived, the fact remains that the circuit is affected by the *positive* voltage V_G at the beginning of the new half-cycle in which it is again supplied with power. Because of this 'false signal', C_2 is charged up too fast, causing premature triggering. This can lead to a hunting effect, in which the motor keeps starting up for a moment and then stopping.

The circuit FW serves in the first place to prevent C_2 from being charged up during the flywheel effect, thus avoiding premature triggering. As long as V_G is positive, the Zener diode Z_2 and the diode D_2 are not involved (Z_2 conducts, D_5 blocks), and the voltage divider formed by the resistors R_6 and R_7 (with $R_8 + R_9$ in parallel with R_6) keeps the base-to-emitter voltage of the transistor Tr positive. The transistor is then conducting, so that C_2 cannot be charged.

Not only are the complications of the flywheel effect eliminated in this way, but the effect itself and the circuit FW also serve to reduce the starting current, so that the motor and control circuit do not become too hot during the running-up period. This is because the beginning of the charging of C_2, and hence the triggering point, are moved up during the time that V_G is positive.

This also explains the nature of the initial part of the current-speed characteristic (see *fig. 7*) for the system formed by the motor plus the electronic circuit when S_2 is in position *1*: when the speed rises from zero, the flywheel effect decreases (the positive flywheel peak is

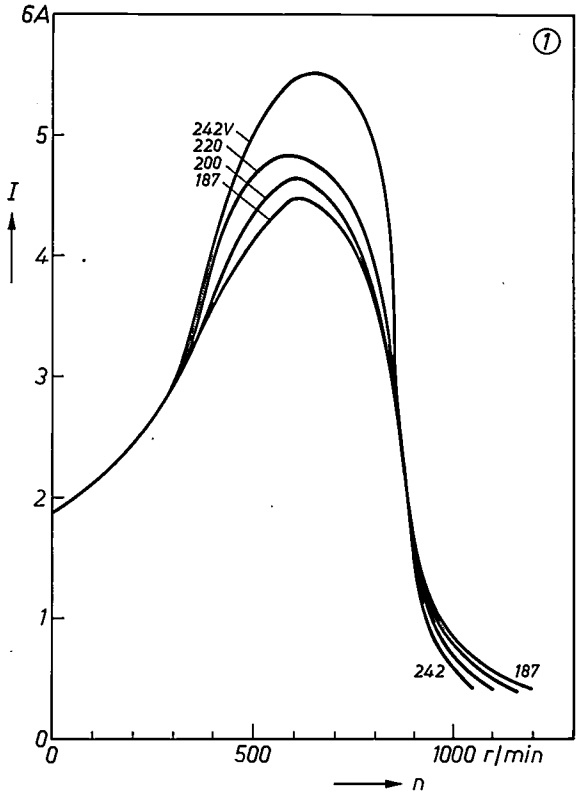


Fig. 7. Current-speed curves of the system of the motor plus the electronic circuit for washing (switch S_2 in position *1*), for various values of the supply voltage. These curves, like those in the following figures, also give the shape of the associated torque-speed characteristics.

swamped in the negative speed voltage), the triggering point is advanced and the motor receives more current. After a while this effect is no longer dominant, and the curves become the normal characteristics for a d.c. motor, but very steep and close together for the various voltages because of the stabilization mechanism. Fig. 7 also gives the shape of the appropriate *torque*-speed characteristics (see eq. 2). The same applies to the later figures (8, 9 and 10).

So far we have not considered the effect of B via the branch $R_{11}R_{10}R_9$ on the base of the transistor. Nevertheless the picture outlined gives a broadly correct description of the operation of the system in position *1*. In position *3*, however, the coupling between B and H is very important, at any rate as long as G is positive or not too strongly negative: as soon as C_2 starts to become charged, H goes positive, the transistor conducts and the charging process is slowed down because charge can leak away through R_{12}. This retards the triggering point to such an extent that the motor no longer receives enough current for starting and remains stationary. In this way the safeguard mentioned earlier against moving straight to 'spin' from the start is effected.

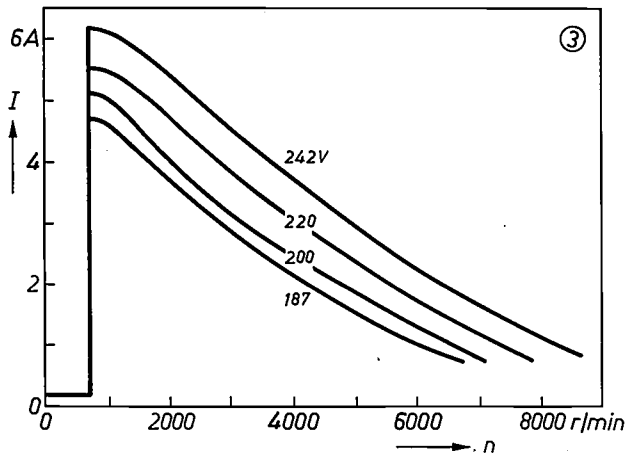


Fig. 8. Current-speed curves for spin-drying (S_2 in position *3*).

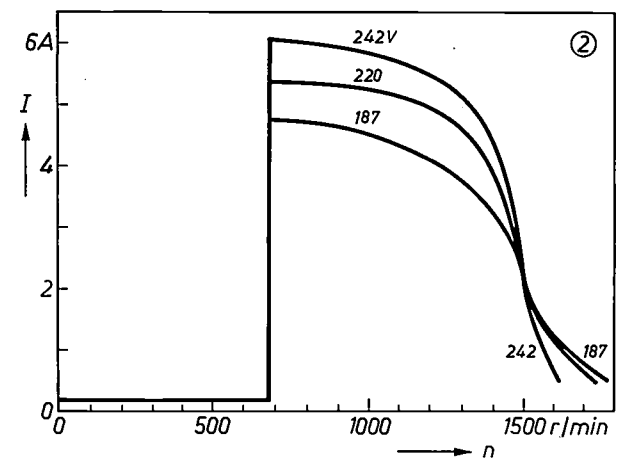


Fig. 9. Current-speed curves for draining (S_2 in position *2*).

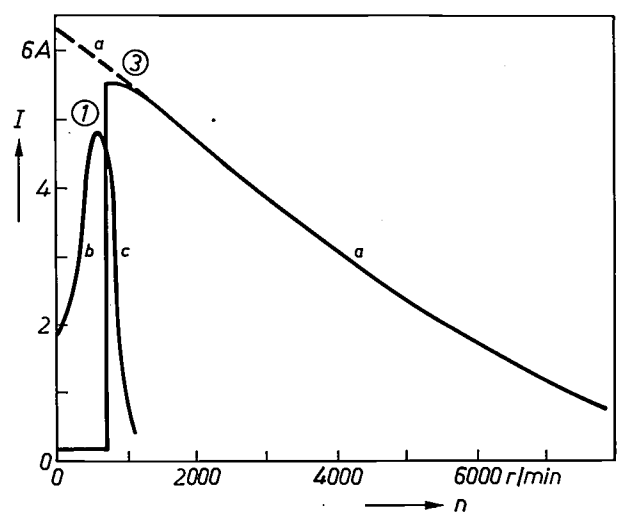


Fig. 10. Combined curves for 220 V from fig. 7 and fig. 8. The branch *a* is approximately the same as the curve for the motor without the electronics. The starting current (*b*) in the 'wash' position was obtained by means of the circuit *FW* in fig. 5. The steepness of curve *c* shows the effect of the speed stabilization in this position.

In position *3* the motor can, however, take power if it is running at a sufficiently high speed. The speed must then be such that $V_b - V_G$ is *greater* than the Zener voltage of the diode Z_2. This diode then conducts and the branch $R_7/D_5Z_2R_6$ mainly determines the voltage V_H, which comes out negative. The coupling between B and H can then no longer prevent the charging of C_2. The diode D_5 protects the base of the transistor from an excessively negative voltage. If $V_b - V_G$ is positive but *smaller* than the Zener voltage of Z_2, then Z_2 may be regarded as an infinite resistance. The coupling between H and B — and consequently the prevention of the charging of C_2 — are then fully effective. Thus, the Zener voltage of Z_2 determines the minimum speed at which the motor can take up power. At a Zener voltage of 7.5 V this speed is 700 rev/min for our motor. We thus arrive at the current-speed curves shown in *fig. 8*, with a threshold at 700 rev/min, and then closely resembling those of the motor without electronics (no speed stabilization).

We have not so far said anything about position *2* (drain). This represents an intermediate situation: the speed has to be stabilized as in position *1*, but the motor must also be prevented, as in position *3*, from running up after stopping. These functions are fulfilled by the Zener diode Z_3. When G is weakly positive or negative, Z_3 constitutes a high resistance. The situation is then as if S_2 were in position *3*, thus guaranteeing the required safety. If, however, the motor exceeds a particular speed, G goes so strongly negative that the Zener voltage of Z_3 is exceeded and Z_3 becomes conductive. In qualitative terms we then have the situation of position *1*, with speed stabilization. The Zener voltage of Z_3 determines the stabilized speed (1500 rev/min). *Fig. 9* shows the current-speed curves for position *2*.

Finally, *fig. 10* presents an overall picture of the results, combining the current-speed curves for 220 V from figures 7 and 8. Curve *a* is broadly the characteristic that would be found without the electronics, curve *b* is due to the utilization of the flywheel effect, and the steepness of curve *c* represents the speed stabilization.

Summary. A d.c. motor for washing machines, whose speed can be controlled electronically, has been developed at MBLE. This motor operates through a fixed transmission ratio of 16 : 1 to give the drum the required speeds for washing (50 rev/min), draining (90 rev/min) and spin-drying (500 rev/min). For washing and draining the speed is stabilized against voltage or load variations. The electronic circuit prevents the motor from running up from start to the drain or spin-dry speeds, which can only be reached by way of the wash cycle. A series resistor protects the semiconductor devices in the circuit against current peaks, and during the wash cycle it produces a large voltage drop that produces a good form factor for the motor current. The same resistor is used for heating the water. The motor is designed for simplicity of manufacture and a long working life. The stator is a permanent magnet. The pieces of ferroxdure are very wide, and soft-iron pole pieces ensure a smooth distribution of the flux. The rotor has 22 slots and 44 commutator segments.

The motor is used with a slightly modified electronic unit in washing machines that Philips have been marketing for more than a year.

Philips tech. Rev. 34, 170-179, 1974, No. 7

High-speed solid-rotor induction motors

H. G. Lakerveld

Introduction

A two-pole induction motor that runs from the a.c. mains has a speed of slightly less than 50 revolutions per second (i.e. 3000 revolutions per minute) [1]. The speed of motors with four, six or more poles is lower by a corresponding factor. Some applications, however, require much higher speeds; filament-coiling machines require speeds up to 30 000 rev/min, ultracentrifuges for uranium enrichment spin at 60 000 rev/min, and nylon-thread spinning machines require speeds as high as 100 000 rev/min. Speeds as high as this are best obtained with a direct drive from high-speed motors supplied by a 'high-frequency' a.c. supply. Other examples are high-speed hand tools (grinding machines, 25 000 to 60 000 rev/min) and machines for processing diamond dies (80 000 rev/min). In vacuum cleaners, which require speeds of up to 20 000 rev/min, the drive is provided by an a.c. commutator motor.

Apart from the technical requirements of the application, there is another argument for using high-speed motors as the drive, and this is the improvement in the power-to-weight ratio of the motor. The *torque* that can be delivered is limited by the dimensions of the motor and by the scope for dissipating the heat generated; the delivered *power*, however, is determined by the product of torque and speed, and the power limit is only reached when the mechanical construction of the motor does not allow a higher speed. Increasing the power-to-weight ratio by raising the speed is of particular advantage when weight is an important consideration, as it is in aeronautics and space technology and also in some electrical automobiles, as yet in prototype; in many cases the advantages of a high-speed motor are not cancelled out by having to use a reduction gear.

High-speed electric motors can be of either the synchronous or the asynchronous type. In applications where a synchronous motor is required a hysteresis motor or a reluctance motor (with starting cage) is usually suitable; a motor with a permanent-magnet rotor is not so suitable because many magnetic materials cannot withstand very high centrifugal forces. If some variation of speed with load is permissible, the obvious solution is an induction motor. This is the type we shall be concerned with in this article, which describes an investigation in which the objective was an optimum design for a high-speed induction motor with a solid iron rotor [2]. In this sturdy construction the rotor serves as both a magnetic and an electric conductor. The iron losses are considerably reduced when the stator windings are located in the air gap instead of in slots. Various types of coil have been specially designed for mounting in the air gap.

Fig. 1 shows two prototypes, both designed for speeds of between 36 000 and 40 000 rev/min and a power of 300 W. The supply for both models can be provided without too much difficulty in the form of a 'square-wave' voltage of the required frequency, which is easier to generate electronically than a sine wave. The small model could for example have been used as a design basis for a vacuum-cleaner motor smaller than the present commutator motors but delivering the same power.

To design the motors it was necessary to calculate the torque. This was no easy matter, because there is no simple way of determining the behaviour of the currents in the rotor; this can only be done by solving Maxwell's equations for the given configuration of stator and rotor. The best way of finding solutions was to start by neglecting the effects at the rotor ends, and then to add later corrections for these unavoidable end effects on a rotor of finite length. A correction was also needed for the effect of the magnetic saturation of the iron.

A more detailed account of our study now follows, in which the rotor, the stator and the calculations of the field and current distribution in the rotor and of the torque are discussed in turn.

Solid rotor

In designing a high-speed induction motor we had to choose between a squirrel-cage motor and a motor with a solid iron rotor. The important factors here were the shape desired for the torque-speed characteristic, and also whether a squirrel-cage motor would have been strong enough. Typical torque-speed characteristics for both motors are shown in *fig. 2*. The characteristic for the squirrel-cage motor (curve *a*) can be changed by altering the resistance of the cage; the smaller the resistance the steeper the curve near the synchronous speed n_0 and the smaller the starting torque [3].

Ir H. G. Lakerveld is with Philips Research Laboratories, Eindhoven.

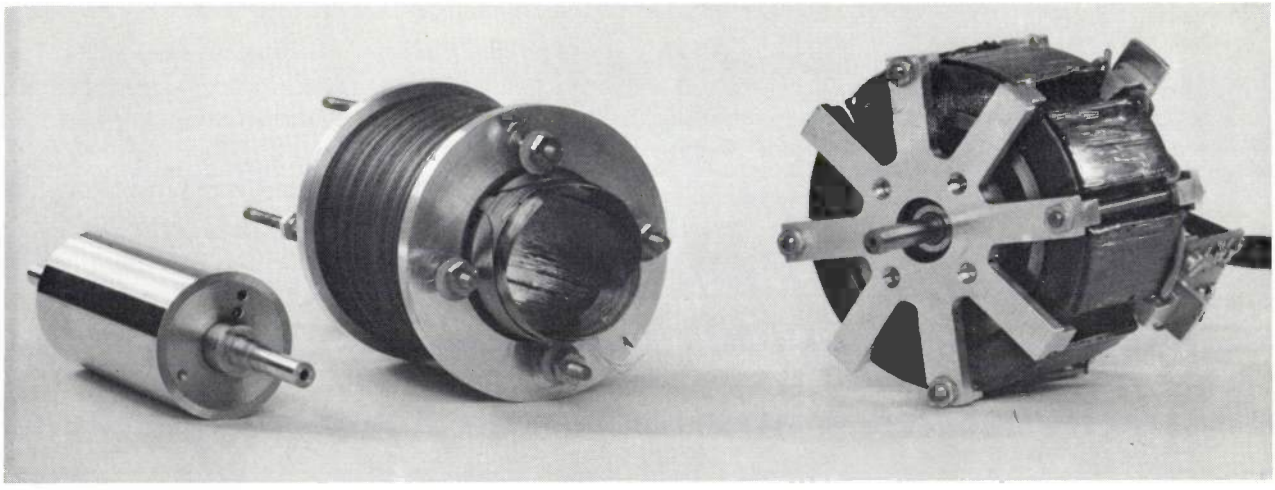


Fig. 1. Two prototypes of high-speed induction motors with a solid iron rotor. The stator coils are not in slots but are situated in the air gap. The motor on the left has thin coils, the one on the right has toroidal coils wound around the stator iron. Both motors were designed for speeds between 36 000 and 40 000 rev/min and a power of about 300 W.

In a solid rotor the currents are forced out towards the surface as the frequency increases. The rotor resistance is therefore a function of the slip. This results in a torque-speed characteristic that is almost flat until the synchronous speed is approached, when it drops steeply to zero. The ratio of the starting torque to the rated torque is therefore better than that of a cage rotor.

To obtain a torque-speed characteristic of this shape with large motors, the rotor is either provided with a double cage — two concentric cages, the outer one with the highest resistance and the lowest leakage inductance — or the bars of the cage are made very deep (skin-effect cage). A double cage is not used for small motors because it takes up too much room. A skin-effect cage is not used in small motors either, because the skin depth is greater than the depth of the bars.

It is also clear that in applications where a high-speed rotor of high mechanical strength is required a squirrel-cage rotor will present more constructional problems than a solid rotor. For this reason and for the reasons mentioned earlier, a solid rotor is to be preferred — even though a squirrel-cage rotor of the same volume delivers a greater pull-out torque (the maximum torque).

A solid rotor also has various incidental advantages. Firstly, since the currents only penetrate into the outermost layer, because of the skin effect, the rotor can take the form of a hollow cylinder, which reduces the moment of inertia. Secondly, the starting current is only about twice as high as the rated current, which is an advantage in the design of the electronic converter that supplies its high-frequency current. Finally, the electrical impedance of the motor does not vary much with the motor speed. Consequently a simple single-phase electronic source can be used for the supply; the auxiliary capacitor included in series with the auxiliary winding of the motor can have the same value for starting and for the rated speed.

A disadvantage of the motor with a solid rotor is its sensitivity to higher harmonics in the stator-field distribution. These 'spatial' harmonics should be distinguished from the time harmonics that appear when the motor is run from a non-sinusoidal voltage supply. Spatial harmonics give rise to slowly rotating stator fields and are the cause of high losses in the rotor. The

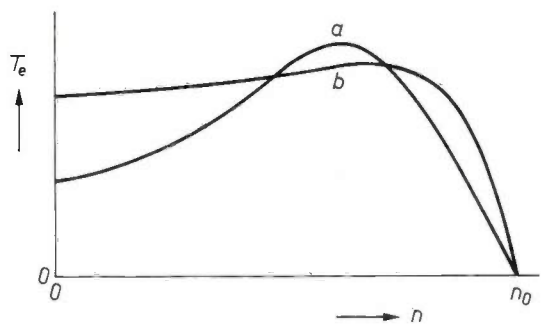


Fig. 2. The torque T_e of a squirrel-cage motor (*a*) and of a motor with solid rotor (*b*), as a function of the speed n. Curve *b* shows a better starting torque. n_0 synchronous speed.

[1] 60 revolutions per second in North America (3600 revolutions per minute).

[2] A more appropriate term would be homogeneous iron, since what is referred to is the absence of copper bars; in any case the rotor is sometimes not solid but hollow. The term solid iron will be used here, however, to be consistent with current usage.

[3] An introductory treatment is given in: E. M. H. Kamerbeek, Electric motors, Philips tech. Rev. **33**, 215-234, 1973 (No. 8/9).

method for 'filtering out' the most undesirable higher harmonics in a squirrel-cage rotor by choosing the appropriate number of rotor bars cannot be used in a solid-rotor motor. This means that when a solid rotor is used it is of paramount importance to have a stator field that is as purely sinusoidal as possible.

Stator without slots

The windings that generate the stator field are usually accommodated in slots in the inside wall of the stator and parallel to the axis. The stator consists of a stack of laminations that can take the form illustrated in *fig. 3a*. The turns are distributed among the slots in such a way as to approximate as closely as possible to the required sinusoidal copper distribution around the circumference, but the approximation is necessarily rather poor. Furthermore, because of the slots the air gap is not completely uniform, but varies periodically around the circumference. When a solid rotor is used this circumferential variation introduces even larger deviations from a sinusoidal stator field than those due to putting the coils in slots. All these variations have a considerable effect (see fig. 7*a*).

A solution to this difficulty can be found by making the bore of the stator smooth instead of slotting it (fig. 3*b*). The stator windings then have to be introduced into the air gap and distributed in the best possible way. Even though they are made as thin as possible, this still inevitably means a considerable widening of the air gap, resulting in a smaller flux density and hence a smaller pull-out torque for the same rotor radius; at the same time the pull-out slip becomes greater. For the same outside diameter of the stator, however, the rotor can be given a larger diameter (fig. 3), which in turn increases the torque.

Apart from this, there are other reasons why a motor with a solid rotor and stator coils in the air gap should have characteristics at least as good as those for a motor with the stator coils in slots — even when the rotor volume is the same. First of all, the effective air gap with a slotted stator is greater than the geometrical air gap, since the flux is concentrated at the stator teeth, where it saturates the iron, making the permeability appreciably lower. The difference in magnitude of the air gap for the two arrangements is therefore not as great as it might seem. Secondly, air-gap coils can take greater current densities since their shape gives a large surface of contact with the air and hence better cooling than for coils in slots. Thirdly, in the case of a motor with a conventional stator some 20% of the calculated fundamental torque must be subtracted for the torque produced by the higher harmonics of the field distribution, which is negative near the nominal speed. There are hardly any higher harmonics in a motor with good air-gap coils. Finally, a stator with a smooth bore has lower iron losses. The flux concentrations at the teeth of a slotted stator contribute substantially to the iron losses because both the eddy-current losses and the hysteresis losses are approximately proportional to the

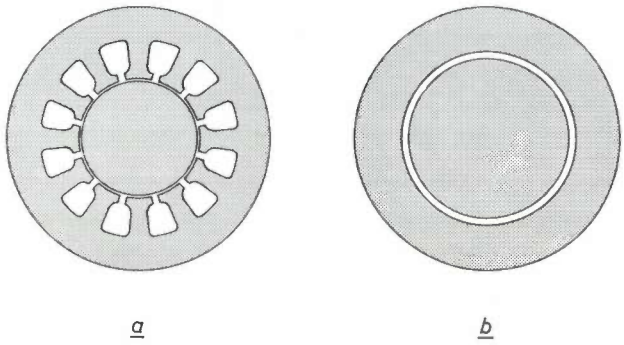


Fig. 3. *a*) A stator with slots for the windings. *b*) A stator with a smooth bore. The windings now have to be accommodated in the air gap.

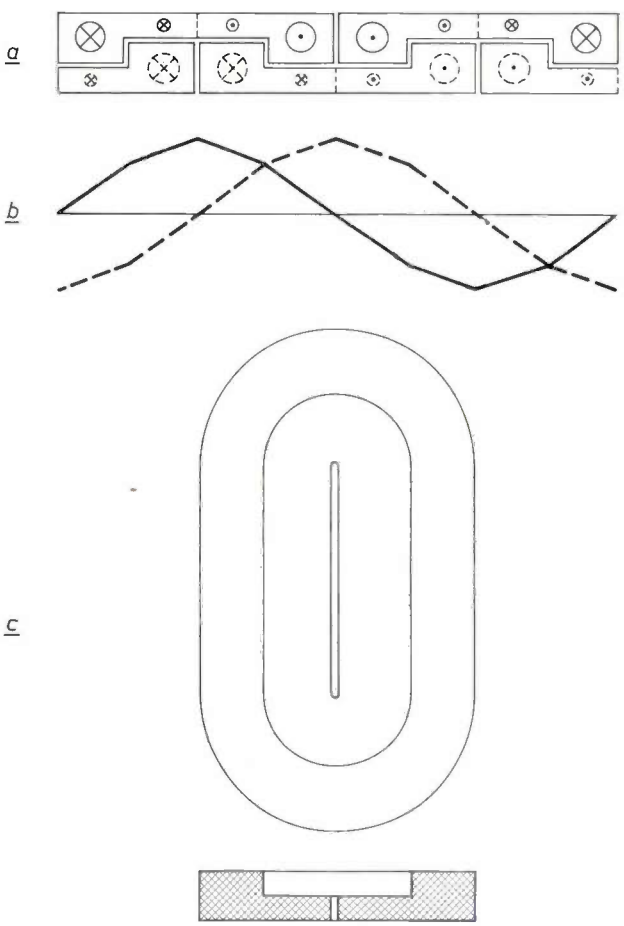


Fig. 4. Air-gap coils that produce an approximately sinusoidal field distribution and at the same time make the maximum use of the space in the air gap. The stator is wound for a two-phase supply. *a*) Cross-section of the (developed) stator coils. The upper coil is for one phase, the lower coil for the other one. *b*) The field pattern for the two phases (solid and dashed curves). *c*) Plan view and cross-section of a coil.

square of the magnetic flux density. In a cylindrical stator stack such flux concentrations are not encountered; furthermore the iron volume is smaller, which also helps to reduce iron losses. This is especially important in high-speed motors, since the iron losses increase approximately as the 1.5th power of the frequency and can reach considerable values at high frequencies. If an upper limit is set for the dissipation from a given stator volume, any reduction in the iron losses allows more heat to be generated in the coils, so that the air-gap coils can carry a higher current, which will produce a higher torque.

Air-gap coils

If it is desired to keep the gap as small as possible when using air-gap coils, and to make the maximum use of the space available for the windings, the coils cannot be arranged to give a purely sinusoidal copper distribution [4]. An approximation must then be made. It appears, however, that a reasonably good sinusoidal field distribution can be obtained with a very simple copper distribution. A distribution of this type is illustrated in *fig. 4a*, which shows a cross-section of the stator coils, developed along a straight line. The motor is assumed to be a two-phase machine (an obvious choice with electronic supply). The solid symbols indicate the sense of the current in the coils of one of the phases, and the dashed symbols indicate the sense of the current in the other. The resultant field distribution can be seen in fig. 4*b*, and fig. 4*c* shows a shape of coil that will give the current distribution illustrated. The thickness of this coil is stepped; if coils of the same thickness everywhere are required, the same current distribution can be obtained with a combination of the two coil shapes illustrated in *fig. 5*.

The large overhang of the coils in figs. 4 and 5 is a disadvantage because a substantial proportion of the energy is uselessly dissipated there, and also because they make the motors unnecessarily long. This disadvantage is not found with the toroidal arrangement shown in *fig. 6*. Here the stator-coil turns are completed outside the stator iron, fig. 6a shows the current directions for a stator with eight coils, each consisting of an inner winding and an outer winding with twice as many turns as the inner one. The solid arrows and symbols indicate the magnetic flux and current directions for one of the windings, and the dashed symbols those of the other. Fig. 6*b* shows a side view and cross-section of a coil, and fig. 6*c* shows the arrangement of the eight coils in the motor (see also fig. 1).

[4] A sinusoidal copper distribution is used in the 'printed' air-gap coils described in the article by E. M. H. Kamerbeek, Torque measurements on induction motors using Hall generators or measuring windings, this issue, page 153. Here, however, the filling factor is very low.

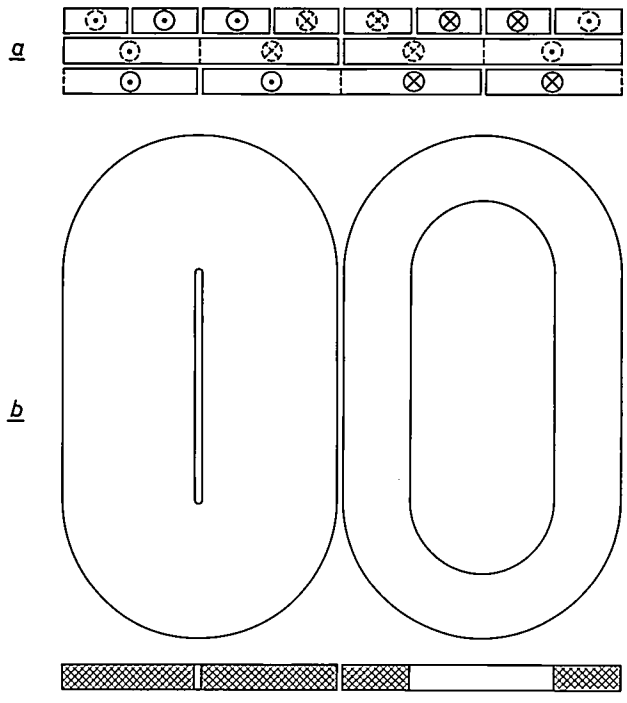


Fig. 5. Another arrangement of thin coils that will produce a field distribution like that in fig. 4*b*. *a*) Cross-section of the stator copper distribution; the solid symbols indicate one phase and the dashed symbols the other. *b*) The two shapes of coil forming the stator windings, seen in plan and in cross-section.

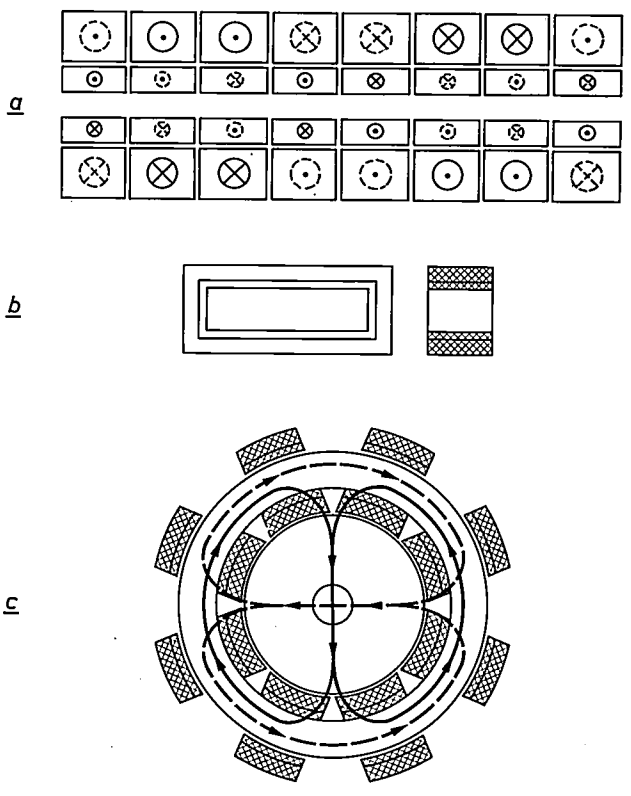
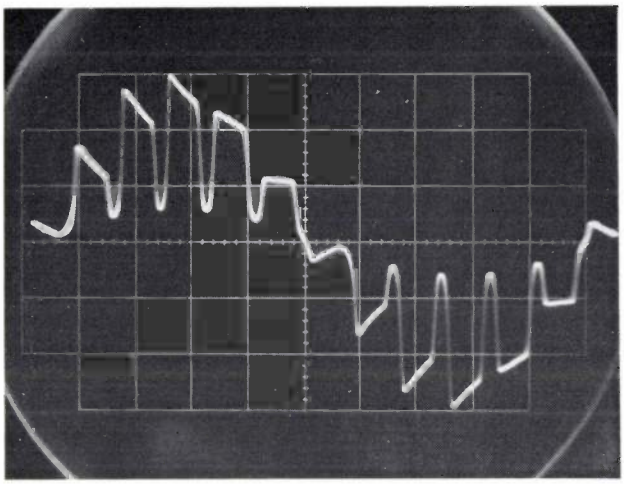
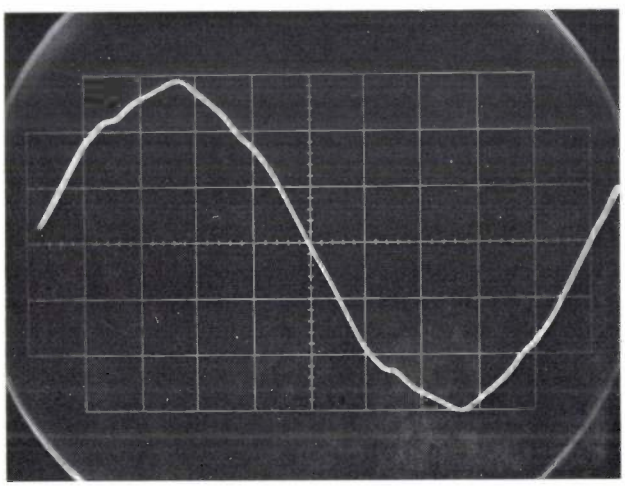


Fig. 6. Toroidal air-gap coils wound on the stator iron. Unlike the coils in figs. 4 and 5, they do not have large overhangs. *a*) Cross-section of the copper distribution. Each of the coils consists of an inner coil and an outer coil with twice as many windings as the inner one. *b*) View of a coil in plan and cross-section. *c*) Cross-section of a motor with eight toroidal coils. The two fields are represented schematically.

The superiority of the field distribution obtained with air-gap coils, even with the simple copper distribution discussed here, can be seen from *fig. 7*. Fig. 7*a* shows the result of a field-distribution measurement in a stator with 12 slots, and fig. 7*b* shows the field distribution when air-gap coils like those in fig. 6 are used. The approximation to a sine curve here is a remarkably good one.



a



b

Fig. 7. Results of measurements of the stator-field distribution (*a*) for a slotted stator, (*b*) for a stator with air-gap coils as in fig. 6. The approximation to a sinusoid is far better in the second case. The measurement was made by exciting one stator phase with d.c. and rotating a search coil with the rotor.

Calculation of the torque

The torque of an electric motor can be calculated if the radial and tangential components of the magnetic field are known at every point of a surface situated in the air gap and enclosing the rotor [5]. In our case the calculation of these components requires the direct solution of Maxwell's equations for the air gap and for the iron of the rotor and the stator. This is possible if we introduce a number of simplifying assumptions. The principal ones are that the stator iron has infinitely high permeability so that the magnetic field-strength in it is zero, and also has zero electrical conductivity. Another assumption is that the solution is independent of the coordinate along the motor shaft, which implies that there are no perturbing effects from the end faces of the rotor (we treat the problem as if these end faces were infinitely far away, so that all the currents in the rotor run parallel to the shaft). Furthermore the permeability of the rotor iron is assumed to be constant, magnetic saturation does not therefore enter the argument. In the next subsection the expressions thus found will be corrected for the end effects of the rotor and for magnetic saturation.

The magnetic fields calculated for this simplified model depend on the slip because the skin depth is different at different speeds. These fields are used for calculating the torque; the expression found is complicated, and will not be given here. A simplification is possible, however, if the values of the slip are not too small. This becomes apparent when we derive the elements of the equivalent circuit of the motor from our knowledge of the magnetic field. This equivalent circuit [3] is shown in *fig. 8*; the form in fig. 8*b* can be derived from fig. 8*a*, and the accented quantities are known as referred quantities. We shall use them several times in the following calculations. In our case the rotor resistance R_r and the leakage inductance of the rotor $L_{r\sigma}$, unlike those of the squirrel-cage motor, are functions of the motor speed and hence of the slip s. If the slip is greater than a minimum value s_{min}, it is found that both R_r and $L_{r\sigma}$ are simply related to s. This can be seen in *fig. 9*, where it is shown that R_r is proportional and $L_{r\sigma}$ inversely proportional to $\sqrt{s}$. The proportionality constant $R_r/\sqrt{s}$ will be referred to as r_r and $R_r'/\sqrt{s}$ as r_r'.

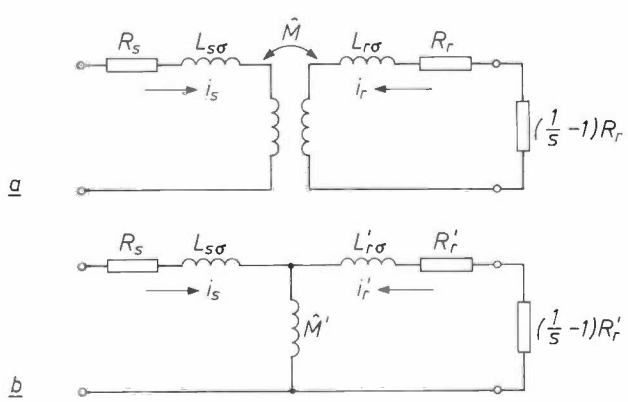


Fig. 8. Equivalent circuit for the induction motor. *a*) Non-referred form. R_s and R_r are the resistances, $L_{s\sigma}$ and $L_{r\sigma}$ the leakage inductances of stator and rotor. The values of R_r and $L_{r\sigma}$ with the solid rotor depend on the slip s. $\hat{M}$ is the mutual inductance between stator and rotor. The resistance $(1/s - 1)R_r$ represents the mechanical load; the total resistance in the rotor circuit is R_r/s. *b*) Referred form. The values assigned to the referred quantities (accented) are such that the same characteristics are measured at the input terminals as in (*a*).

The complicated expressions that we derive for R_r' and $L_{r\sigma}'$ can then be simplified, as can also the expression for the torque.

To derive the rotor resistance R_r' and the various inductances we use another model of the infinitely long rotor. We assume in this new model that the currents flow in a thin sinusoidal copper layer applied to the surface of the rotor; the rotor iron is assumed to be non-conducting and to have the same permeability as in the previous model. In this way we can obtain a simple definition for one value of the rotor resistance. This modification must not cause any change in the rotor dissipation, the magnetic field in the air gap or the total magnetic field energy of the system. The currents in the hypothetical copper layer can be derived from the calculated magnetic field. Since for a given torque T_e and an angular frequency ω of the stator currents the total dissipation P_r in the rotor is known from the relation

$$P_r = s\omega T_e,$$

we can now calculate the rotor resistance.

The distributed inductance of this rotor model with its copper layer is very low. The reason is that the sinusoidal stator and rotor windings in this model are both assumed to be in the air gap, so that the coupling is virtually ideal. To allow for the magnetic field energy present in the real motor we must include a hypothetical leakage inductance $L_{r\sigma}$ in the rotor circuit, compared with which the leakage inductance of the model rotor is negligible. The $L_{r\sigma}$ plotted in fig. 9 is this hypothetical value. Calculation shows that $R_r' = s\omega L_{r\sigma}'$ for $s > s_{min}$.

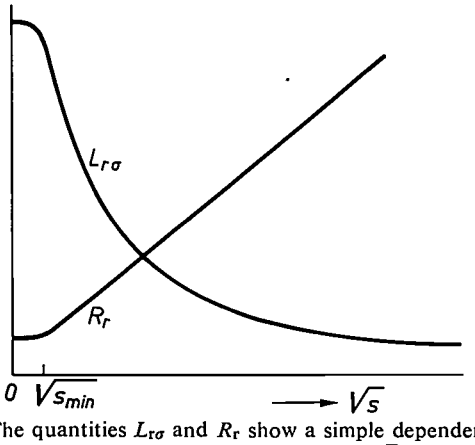


Fig. 9. The quantities $L_{r\sigma}$ and R_r show a simple dependence on the slip values $s > s_{min}$; in such cases $L_{r\sigma} \propto 1/\sqrt{s}$ and $R_r \propto \sqrt{s}$.

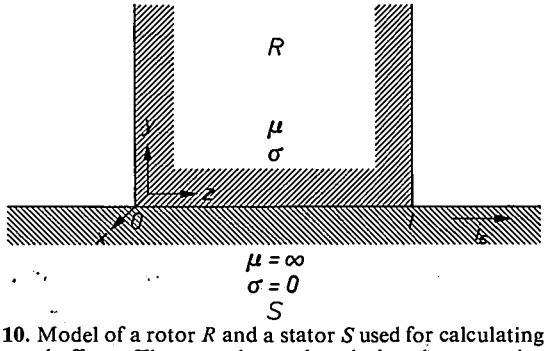


Fig. 10. Model of a rotor R and a stator S used for calculating the rotor end effects. The rotor has a length l and moves in the x-direction relative to the stator.

The forms which the referred values R_r' and $L_{r\sigma}'$ take when $s > s_{min}$ are:

$$R_r' = \pi a l \hat{z}_s^2 \sqrt{s\omega\mu/2\sigma},$$
$$L_{r\sigma}' = \pi a l \hat{z}_s^2 \sqrt{\mu/2s\omega\sigma}.$$

Here a is the radius, l the length, μ the permeability and σ the conductivity of the rotor; $\hat{z}_s$ is the peak value of the copper-distribution function for the stator.

The mutual inductance between stator and rotor ($\hat{M}$ in the equivalent circuit) can be calculated in the usual way from the dimensions and other characteristics of the model. At a given amplitude $\hat{\imath}_s$ of the a.c. currents in the stator the torque is now given by:

$$T_e = \hat{\imath}_s^2 \hat{M}/(\sqrt{s/s_{max}} + \sqrt{s_{max}/s} + \sqrt{2})\sqrt{2},$$

and the pull-out torque by:

$$T_{max} = \hat{\imath}_s^2 \hat{M}/(2 + \sqrt{2})\sqrt{2} = \hat{\imath}_s^2 \hat{M}/4.83,$$

while

$$s_{max} = 2r_r'^2/(\omega\hat{M})^2$$

is the slip at which the motor develops the pull-out torque.

Effect of finite rotor length

As mentioned earlier, the results found need to be corrected because in a real rotor the currents are not axial everywhere but flow along closed paths, so that particularly near the end faces there are tangential and radial current components, which affect the operation of the motor [6].

To establish the magnitude of the correction these current components and the magnetic fields associated with them must be known. We are able to determine these approximately on the basis of a model different from those used previously. From the currents and fields we are then able to calculate the equivalent resistance and leakage inductance of the rotor as shown in the equivalent circuit in fig. 8. By comparing these quantities with the corresponding ones for a rotor with no end effects we arrive at a correction factor.

The model used here is shown in *fig. 10*. The stator is assumed to be infinitely long, whereas the rotor has

[5] See E. M. H. Kamerbeek, Torque measurements on induction motors using Hall generators or measuring windings, this issue, page 153.

[6] H. Yee, Effects of finite length in solid-rotor induction machines, Proc. IEE **118**, 1025-1033, 1971.
H. Yee and T. Wilson, Saturation and finite-length effects in solid-rotor induction machines, Proc. IEE **119**, 877-882, 1972.

a finite length l along the z-axis. The air gap is neglected [7]. The rotor moves along the x-axis at a velocity v with respect to the stator field. In this configuration a solution for Maxwell's equations can be found that gives the magnitude and direction of the magnetic flux density and the current density at every point of the rotor.

the end faces than in the middle of the rotor, where the values found are about the same as those for an infinitely long rotor. Consequently at the ends of the rotor the product of B_y and the local tangential field-strength H_ϕ produced by the stator is greater than in the infinitely long rotor, and this product $B_y H_\phi$ (the Maxwell stress [5]) is a measure of the torque acting on the

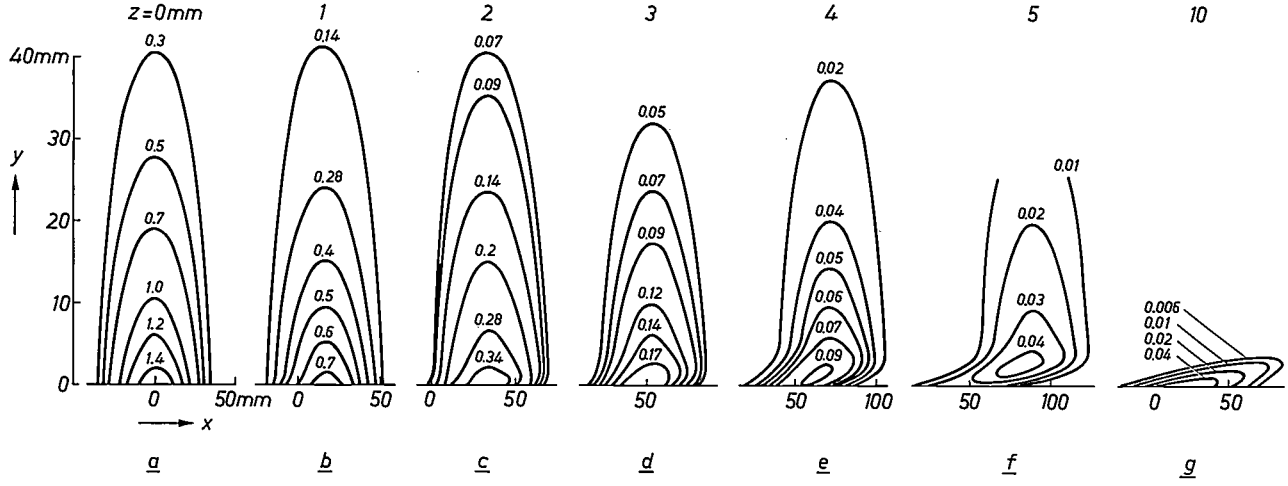


Fig. 11. Calculation of the magnetic field in the rotor of fig. 10. The figures *a-f* show the lines of force in the planes for which $z = 0, 1, \ldots, 5$ mm; fig. *g* relates to the plane $z = 10$ mm. All figures assume that the rotor moves in the x-direction at a velocity of 20 m/s relative to the stator field. The skin depth y in millimetres is plotted vertically, and the coordinate x measured from a zero point in the stator field pattern is plotted horizontally. The parameter for the curves gives the magnetic flux (per metre length along the z-axis) lying outside the curve. The difference between two neighbouring parameter values thus indicates the flux (per metre length along the z-axis) between the associated curves.

Some idea of the magnetic fields at and near the end faces can be obtained from *fig. 11*, which shows 'snapshots' of the lines of force in planes parallel to the end of the rotor and at various distances z from it; the slip velocity v is 20 m/s. The parameter for the curves is the magnetic flux (per metre length along the z-axis) lying outside a given curve; the unit is Wb/m. The difference between two neighbouring parameter values therefore gives the flux (per metre length along the z-axis) between the two curves. The numerical values are based on the assumption that the conductivity of the rotor iron is 5.10^6 A/Vm. At the end face ($z = 0$, fig. 11*a*) the field is equal to that in air and is symmetrical with respect to the abscissa 0, which corresponds to a null in the stator field. Even at a distance $z = 1$ mm from the end face (fig. 11*b*) the magnetic field can be seen to penetrate less deeply into the rotor iron along the y-axis and is also beginning to lag behind the stator field (moving to the left in fig. 11). This tendency increases with increasing distance from the end face (fig. 11*c-g*), and in the middle of the rotor the field is very little different from that in the infinitely long rotor.

We see clearly from these figures that the component B_y at the surface of the rotor is considerably larger near rotor. The field pattern shown in fig. 11 thus indicates that the end effects will lead to a higher torque. It also appears from fig. 11 that these end effects extend over several millimetres along the length of the rotor, and can thus make a significant contribution.

The total resistance of the current paths in the rotor will clearly be strongly affected by the route they take. If the current flows more or less directly across the rotor end face it has to cover a path $\pi/2$ times shorter than if it followed the circumference. To get some idea of the pattern of the currents we determined the current density and current direction in the end face. The lines in *fig. 12* connect places of equal current density; the numbers indicate the current density in A/mm^2, and the arrows the direction of the current in the end face. It can be seen that the current densities at a distance of about 1 cm from the circumference of the rotor have decreased only very slightly; there can be no doubt that the currents do not only travel along the edge but across the end face by a shorter route.

Let us now see, in quantitative terms, what influence the end effects in the rotor have on the rotor parameters. This we can do by calculating the resistance and leakage inductance of the rotor using the model in

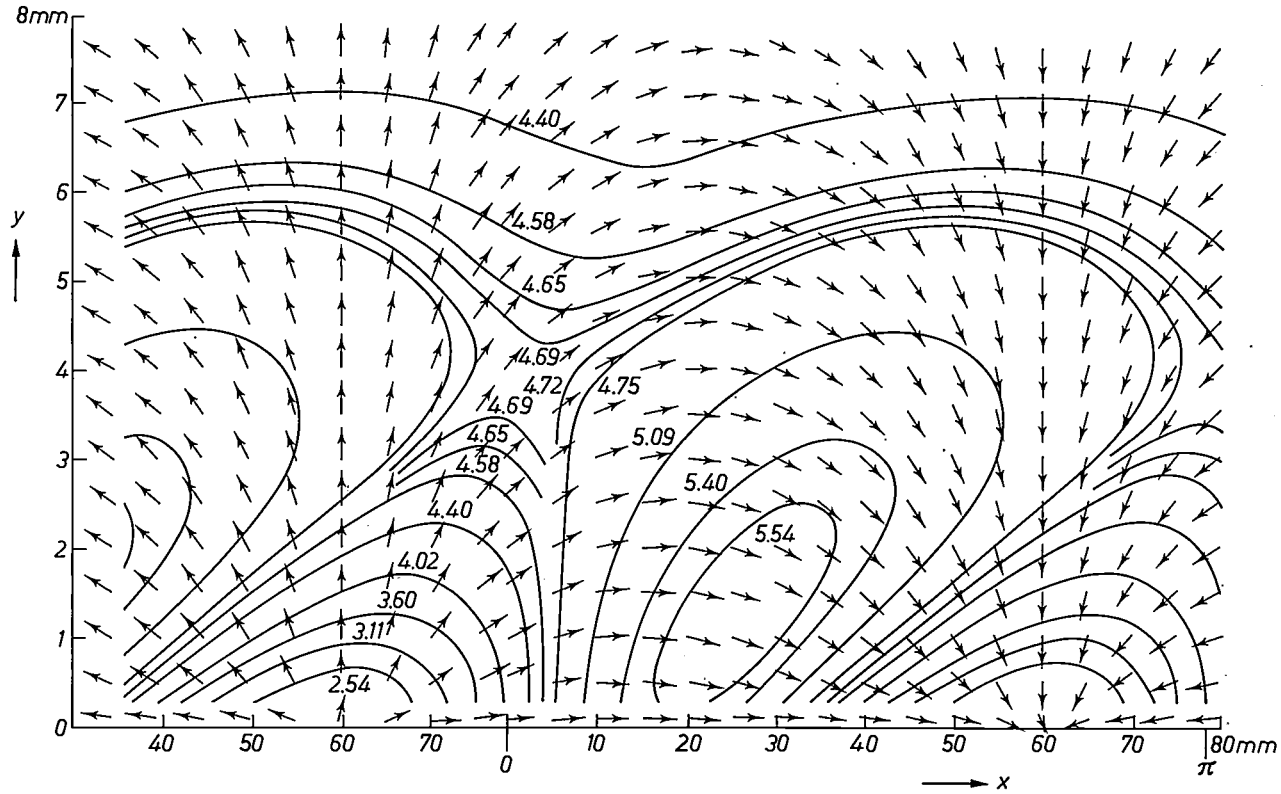


Fig. 12. Calculated currents in the end face of a rotor like the one in fig. 10. The curves connect the points of equal current density; the numbers beside the curves give the current density in A/mm². The arrows indicate the sense of the current component in the x,y-plane; the current component along the z-axis is not shown. It can be seen that the currents do not only flow at the edge of the rotor but extend well over the end face.

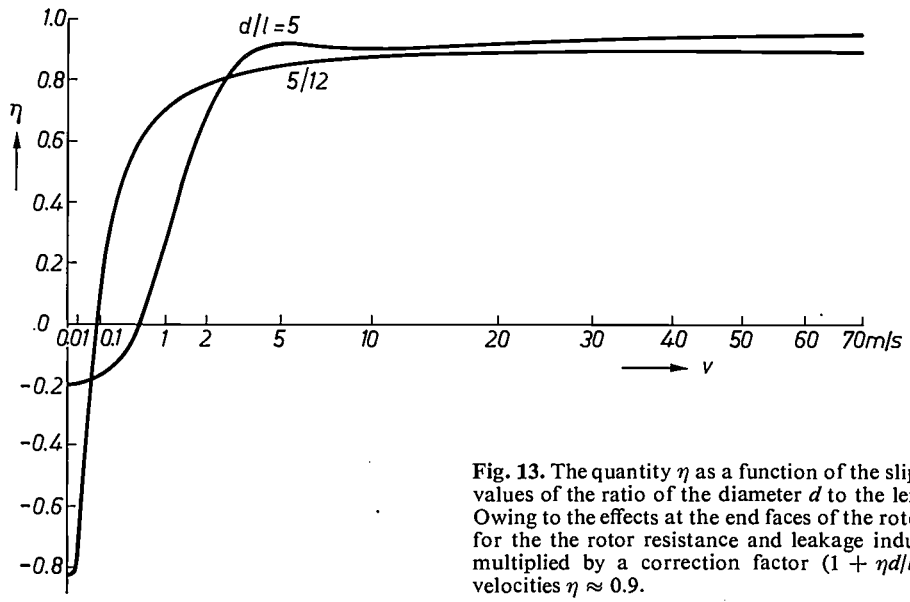


Fig. 13. The quantity η as a function of the slip velocity v for two values of the ratio of the diameter d to the length l of the rotor. Owing to the effects at the end faces of the rotor the values found for the the rotor resistance and leakage inductance have to be multiplied by a correction factor $(1 + \eta d/l)$. At greater slip velocities $\eta \approx 0.9$.

fig. 10. When we compare these with the values found for a piece of length l of an infinitely long rotor, we find that both quantities are greater by a factor of $(1 + \eta d/l)$ for a rotor of finite length l and diameter d. Here η is a function of the slip velocity and also of d/l; the value of η is given in *fig. 13* for two values of d/l. We see that, except for very low slip velocities, η is positive and is equal to about 0.9 at higher slip velocities. The pull-out slip s_{max}, which is proportional to the square of the rotor resistance, is greater by a factor of $(1 + \eta d/l)^2$ for a finite rotor length; this implies that the torque-speed characteristic near the synchronous speed is less steep the shorter and thicker the rotor.

[7] The air gap is included in the calculations made by W. P. A. Joosen, Finite-length effect in a solid-rotor motor, Philips Res. Repts. **28**, 485-495, 1973 (No. 5).

Magnetic saturation

It now only remains to correct our results for the magnetic saturation in the rotor iron, which until now has been assumed to be of uniform permeability. When we determine the paths of the magnetic lines of force in a cross-section of the rotor in our original model (in which the end effects were neglected) we obtain results like those shown in *fig. 14*. It can be seen from this figure that the lines of force become denser at the surface of the rotor, and that the density increases as the slip with respect to the stator field increases. This eventually leads to magnetic saturation in the outer layer of the rotor.

If the existing nonlinear relation between permeability and field-strength is introduced into the equations, the calculations become impossibly complicated. Here again we have to make use of a simplified model. It is known that an exact calculation of magnetic fields and current densities can be made for the case of an infinite half-space of ferromagnetic material characterized by a 'rectangular' *B-H* curve as shown in *fig. 15*, at whose surface there is a sinusoidally alternating tangential magnetic field [8]. This model shows some resemblance with the solid rotor in which also, as can be seen in fig. 14, the lines of force soon begin to bend over parallel to the surface.

As in the case of the rotor, we can again define an equivalent resistance and a leakage inductance for the model. The impedances of both are identical in modulus if we assume constant permeability, as we also found in the case of the solid rotor. If we use the *B-H* curve in fig. 15, however, we find that the equivalent resistance is twice as large as the modulus of the impedance of the leakage inductance. If we can now assume that the occurrence of magnetic saturation in the solid rotor leads to the same change, then for the rotor with magnetic saturation, we have $R_r' = 2s\omega L_{r\sigma}'$.

Applying the corrections both for the end effects and for the magnetic saturation we find the following expressions for the referred values of the rotor resistance and leakage inductance:

$$R_{r,corr}' = \tfrac{16}{3}\, al\hat{z}_s^2 \sqrt{s\omega\mu/2\sigma}\,(1 + \eta d/l),$$
$$L_{r\sigma,corr}' = \tfrac{8}{3}\, al\hat{z}_s^2 \sqrt{\mu/2s\omega\sigma}\,(1 + \eta d/l).$$

This yields a different value for the calculated torque:

$$T_e = \hat{\imath}_s^2 \hat{M}/(\sqrt{s/s_{max}} + \sqrt{s_{max}/s} + 1/\sqrt{1.25})\sqrt{1.25}.$$

The pull-out torque is now greater:

$$T_{max} = \hat{\imath}_s^2 \hat{M}/3.24,$$

a value which agrees very well with measurements. Unfortunately this is not the case with the pull-out slip,

$$s_{max} = 1.25\, r_{r,corr}'^2/(\omega\hat{M})^2,$$

which follows from the same assumption ($r_{r,corr}'$ is the corrected value of the proportionality constant r_r' mentioned earlier to allow for the end effects and the magnetic saturation). The reason for this is that $r_{r,corr}'$ is proportional to $\sqrt{\mu}$, and hence s_{max} to μ; now, however, the permeability of the unsaturated rotor iron can no longer be taken into account, and instead an average permeability must be used for the whole rotor. At the surface the rotor iron is saturated; the permeability there (the saturation induction divided by the tangential field-strength at the surface) is therefore smaller than in the bulk of the rotor. It appears empirically that a

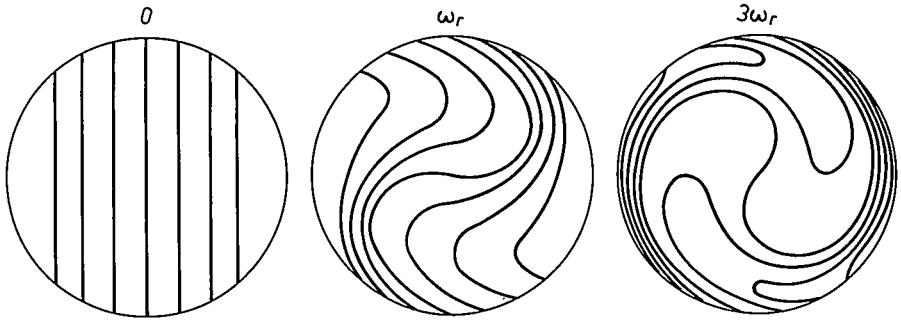


Fig. 14. Lines of force in a cross-section of the rotor for different frequencies of the a.c. current in the rotor. The higher the frequency the more the lines of force concentrate at the surface of the rotor; the outer layer of the rotor finally becomes magnetically saturated, which affects the performance of the motor.

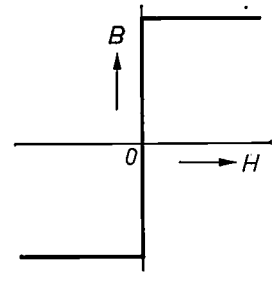


Fig. 15. The 'rectangular' *B-H* curve used for an approximate calculation of the effects of magnetic saturation in the rotor.

[8] W. MacLean, Theory of strong electromagnetic waves in massive iron, J. appl. Phys. 25, 1267-1270, 1954.

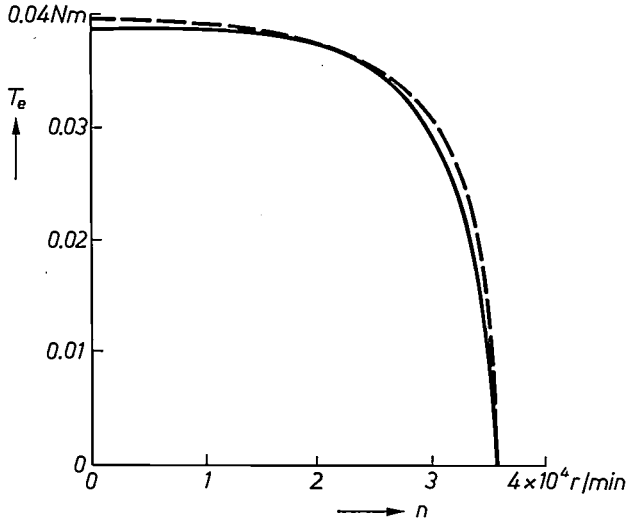


Fig. 16. Measured and calculated torque-speed characteristic of a high-speed induction motor with a solid rotor; the dashed curve is the calculated characteristic. The approximate description of the end effects in the rotor and the magnetic saturation of the rotor iron is accurate enough to bring the calculated characteristic close to the measured curve.

value of twice this surface permeability gives a satisfactory average value. When this is used in calculating the pull-out slip a theoretical torque-speed characteristic is obtained that agrees well with the results of the measurements (*fig. 16*).

Summary. Several prototype induction motors have been developed for speeds of up to 40 000 rev/min. Square-wave voltages at frequencies up to 700 Hz are used for the supply. The iron losses have been reduced and the stator-field pattern improved by making the stator windings in the form of thin coils in the air gap rather than putting them in slots. The motors have a 'solid iron' rotor without copper bars; the rotor currents give an increasing skin effect with increasing slip, which increases the starting torque in relation to the pull-out torque. The torque-speed characteristic is calculated with the aid of simplified models; it is necessary to go back to Maxwell's equations and to solve these for the models. This is done in steps; an infinitely long rotor of homogeneous permeability is first assumed, and corrections are then made to allow for rotor end effects and for the magnetic saturation at the surface.

Philips tech. Rev. 34, 180-189, 1974, No. 7

Supply-voltage speed control for capacitor motors

K. Rennicke

Introduction

One of the most widely used electric motors is the squirrel-cage-rotor machine; it owes its popularity to its simple and sturdy construction without commutator or slip rings and brushes [1]. A drawback, however, is that its speed is difficult to control. The synchronous speed is determined by the mains frequency and by the number of poles of the stator winding. To control the speed of the motor by changing the synchronous speed, it is then necessary to switch to a different number of poles or to alter the frequency of the supply voltage. The first method, pole-switching, is not suitable for controlling the speed of a drive, because it can only change the speed in large steps; the second method, using a variable frequency converter, does provide a technically elegant solution but is rather expensive.

We shall be concerned here with a third method sometimes used to control the speed of squirrel-cage-rotor machines, which is to control the motor torque by altering the stator voltage. This allows continuous speed control to be obtained without changing the synchronous speed. Since this method of control generally involves a lower efficiency, drives controlled in this way are mainly confined to applications where the motor is only switched on for short intervals, and therefore does not become too hot.

Earlier methods of stator-voltage control used thyratrons, thermionic valves, servo-controlled transformers or magnetic amplifiers (chokes with a d.c. premagnetization that saturates the iron core to a greater or lesser degree and so affects the inductance). Nowadays only electronic elements like thyristors or bipolar triode-thyristors (triacs) are used. These are less expensive, lighter in weight and often permit faster control.

Speed control using thyristors or triacs is effected by phase control of the stator voltage, i.e. by switching it on for only a fraction of a half-cycle (*fig. 1*). The magnitude of this fraction is expressed by the conduction angle α ($0° \leqslant \alpha \leqslant 180°$).

The speed control discussed in this article extends over all four quadrants *I* to *IV* of the torque-speed characteristic, but only up to the synchronous speed n_0 (see *fig. 2*). The region above the synchronous speed is of no interest in four-quadrant control, since no driving torque can be developed in that region.

Dr K. Rennicke is with Philips Forschungslaboratorium Hamburg GmbH, Hamburg, Germany.

Control of single-phase squirrel-cage armature motors with an auxiliary capacitor

The phase control of three-phase motors connected to a three-phase mains requires semiconductor devices for each phase. For low-power applications the electronics can soon cost more than the motor itself. The control circuit must therefore be kept as simple as possible. An obvious means of doing this is to use a single-phase instead of a three-phase supply and control; the motor then has two windings, one of which — the auxiliary phase — is fed via an extra impedance, usually a capacitor.

Speed-controlled capacitor motors of this type are widely used in single-quadrant and low-power applications (less than 70 W), for example as servomotors. Since they can be fed from a single-phase supply, they are of greater general use than three-phase controlled motors.

As long as the rated power of the capacitor motors is less than 70 W their behaviour presents no problems. Even at higher powers (up to 800 W) the capacitor motor has distinct advantages, but in four-quadrant use difficulties arise with regard to reversibility, uniformity in speed and stability. In this article we shall attempt to analyse these difficulties. Computer calculations of

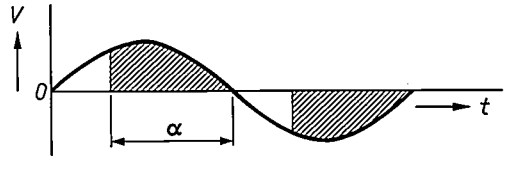


Fig. 1. Phase control. The a.c. voltage $V(t)$ is only switched on for a fraction of each half-cycle (shown by hatching). α conduction angle.

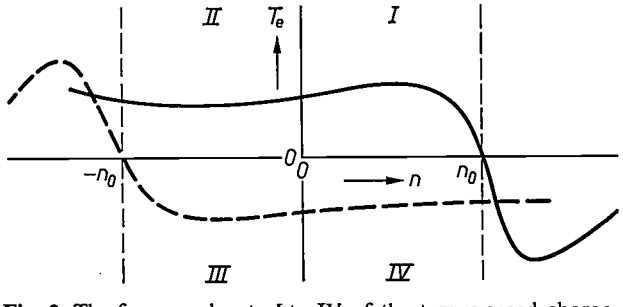


Fig. 2. The four quadrants *I* to IV of the torque-speed characteristic of an induction motor. T_e torque. n speed. n_0 synchronous speed. When the sense of rotation of the stator field is reversed the machine goes from one curve to the other. In quadrants *I* and *III* the machine operates as a motor, in quadrants *II* and *IV* as a brake.

torque-speed characteristics are presented for various values of the capacitor, which permit the most suitable capacitance values to be chosen for reversibility and uniform speed. There is only one speed at which the capacitor 'balances' the motor (i.e. makes it electrically completely symmetrical); at other speeds the ordinary rotary field in the stator is opposed by a field rotating in the opposite direction, which causes torque pulsations at a frequency of 100 Hz. These pulsations, which affect the constancy of speed, have also been calculated.

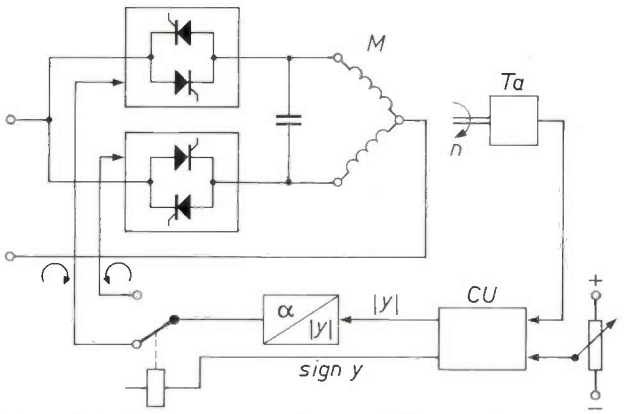


Fig. 3. Circuit for controlling the speed of a capacitor motor M in four quadrants. A tachogenerator Ta whose output voltage is proportional to the speed n is coupled to the motor. The control unit CU comprises a comparator in which this voltage is compared with a set reference value. The result of the comparison is a control signal y, whose magnitude $|y|$ determines the instant at which the thyristors are triggered and thus determines the conduction angle α; sign y controls a relay that determines which of the two motor phases will be connected to the mains voltage; this in turn determines the sense of the motor torque.

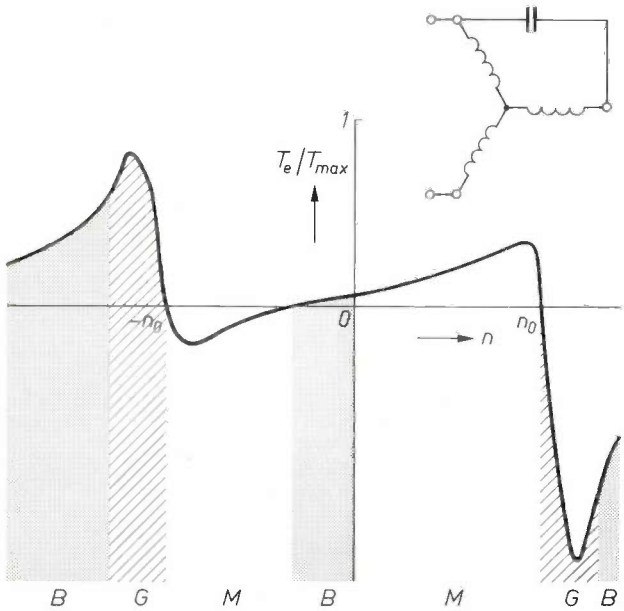


Fig. 4. Torque-speed characteristic of a three-phase machine with a capacitor. The speed ranges in which the machine operates as a motor (M), as a generator (G) or as a brake (B) are indicated. Unlike the case with a symmetrical supply, the machine operates in part of the speed range between $-n_0$ and 0 not as a brake but as a motor. The torque T_e is plotted in its relation to the pull-out torque T_{max}; the capacitor has a reactance $X_C = 0.957\,X$ (X = reactance of unloaded motor).

The stable behaviour of the motor can be upset by the presence of the resonant circuit formed by the capacitor and the inductance of the motor. This can give rise to spontaneous oscillations, particularly at small conduction angles, where the damping due to the low internal resistance of the mains is largely eliminated. An additional resistance is then required in parallel with the motor to provide damping. An analysis of these instabilities is presented here, and a power limit below which the motor can be operated without a parallel resistance is derived from practical data.

Fig. 3 shows a diagram of a simple circuit for controlling the speed of a capacitor motor. A two-phase machine has a capacitor connected across its terminals. The mains voltage V is applied via either the upper or the lower pair of parallel-opposed thyristors, depending on whether the torque is to be positive or negative. Coupled to the motor is a tachogenerator Ta, which generates a d.c. voltage proportional to the speed. This d.c. voltage is compared with a set value in a comparator. The result is a control quantity y, which is resolved into $|y|$ and sign y. The modulus $|y|$ is expressed by the magnitude of a d.c. voltage that controls the conduction angle α via the $|y|/\alpha$ converter. The sign of y is used to control a switch that reverses the sense of rotation of the rotating field.

Apart from the advantages of a smaller number of thyristors and simpler electronics, the four-quadrant control of a capacitor motor as illustrated in fig. 3 also gives a better $\cos\phi$ than a three-phase controlled induction motor.

Operation as a motor, a generator or a brake

An asynchronous motor connected to the symmetrical three-phase mains behaves in the various speed ranges indicated in fig. 2 in different ways. It operates

as a motor	in the range	$0 < n < n_0$,
as a generator	in the range	$n_0 < n < n_1$ [2],
as a brake	in the range	$-\infty < n < 0$ and
		$n_1 < n < \infty$.

Here 'operates as a brake' means that the machine receives energy from the mechanical load and also from the electrical source.

With the capacitor motor these states alternate with one another more frequently — how frequently depends on the value of the capacitor. *Fig. 4* shows the static torque-speed characteristic of a three-phase machine with capacitor. 'Static' here means that the

[1] See the introductory article by E. M. H. Kamerbeek, Electric motors, Philips tech. Rev. 33, 215-234, 1973 (No. 8/9).

[2] The speed n_1, which depends on the resistances and reactances of the machine and on the mains frequency, is not in practice reached.

average torque is plotted, so that the curve does not show the 100 Hz pulsations. The curve relates to a capacitor of reactance $X_C = 0.957\,X$ (X is the zero-load reactance measured at the motor terminals). The ranges are indicated in which the machine operates as a motor (M), as a generator (G) or as a brake (B). It can be seen that in the speed range $-n_0 < n < -0.3\,n_0$ the machine operates as a motor and not, like a symmetrically fed three-phase machine, as a brake.

With a four-quadrant drive the speed control covers the range $-n_0 < n < n_0$. The operation of the machine as a motor in the interval $-n_0 < n < -0.3\,n_0$ makes it impossible, however, to use the machine in four quadrants, since the control in this case cannot establish a positive braking torque.

This brings us to the question of the measures needed to keep the torque-speed characteristic as flat as possible in the control range $-n_0 < n < n_0$, or at least to prevent it from cutting the horizontal axis. The ideal would be to give the capacitor motor a characteristic similar to that of a three-phase motor fed from the three-phase mains. This ideal can be approximated by proper electrical balancing of the capacitor motor.

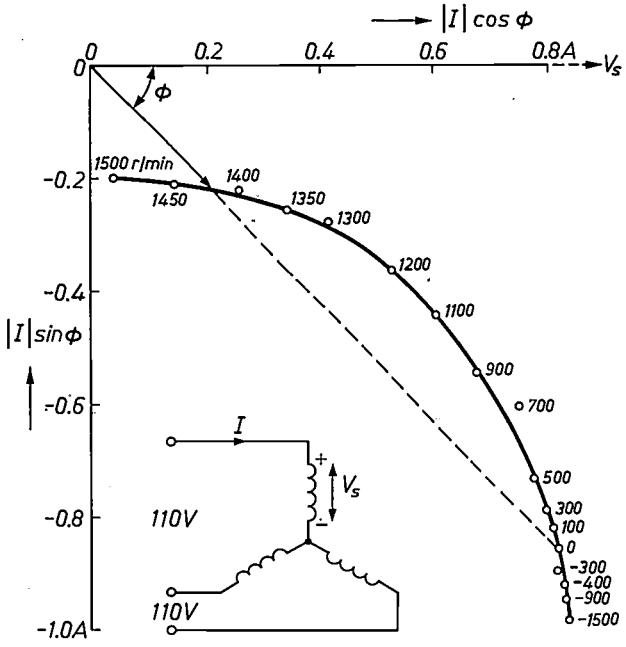


Fig. 5. Magnitude and phase angle ϕ of the input current I of a symmetrically fed 370 W three-phase motor. The figures shown along the curve refer to the speed n. V_s terminal voltage with respect to the star point. Near the nominal speed the current has a phase angle of the same magnitude as at zero and negative speed.

Balancing a capacitor motor

The problem of balancing a capacitor motor electrically is encountered not only in our case but in all cases where a capacitor motor is used, without speed control and often for heavy duty, because there is only a single-phase supply available. The motor then usually has a single nominal speed.

As was noted earlier, with an appropriate choice of the capacitor and other added reactances and resistances, a motor can be balanced for a given speed. The characteristic feature of a balanced motor is that the air gap contains only one field wave rotating at the synchronous speed. In an unbalanced motor the field has two components rotating in opposite directions. The component rotating with the rotor generates a motor or generator torque, while the component rotating in the opposite sense generates a braking torque. The superposition of the two components leads to the characteristic shown in fig. 4. The field rotating in the opposite sense reduces the motor torque and causes extra losses; the higher the motor power the more difficult it is to dissipate the heat developed.

If the motor is to be balanced simply with a capacitor, using no other devices such as an autotransformer, there are two speeds at the most at which this can be done exactly. With a three-phase induction motor the phase angle ϕ between the current and the voltage must then be 60° for each phase; in the case of a two-phase induction motor in which both phases have the same

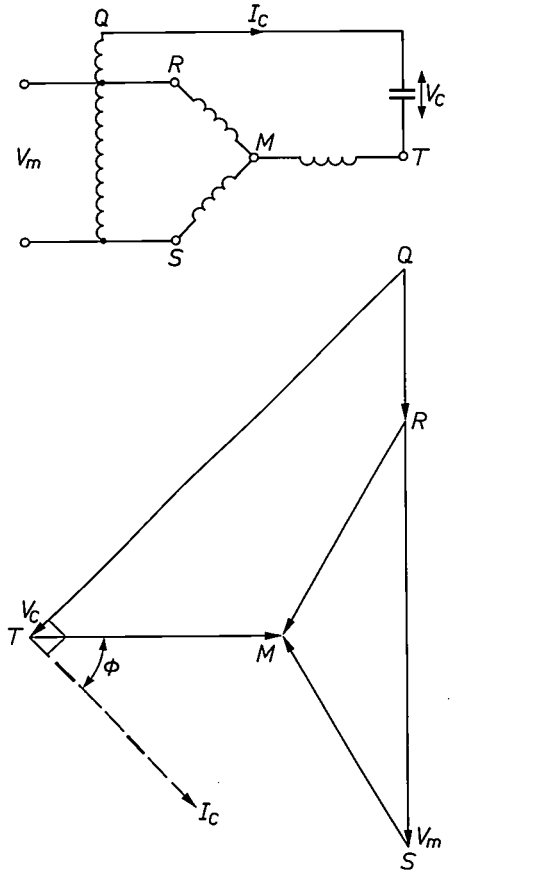


Fig. 6. Circuit and vector diagram for a capacitor motor with autotransformer. V_m mains voltage. V_C voltage across the capacitor C. I_C current through the capacitor (and through the motor phase TM). The vectors V_C and I_C are always perpendicular to one another; if ϕ becomes 60°, then Q coincides with R and the autotransformer can be dispensed with.

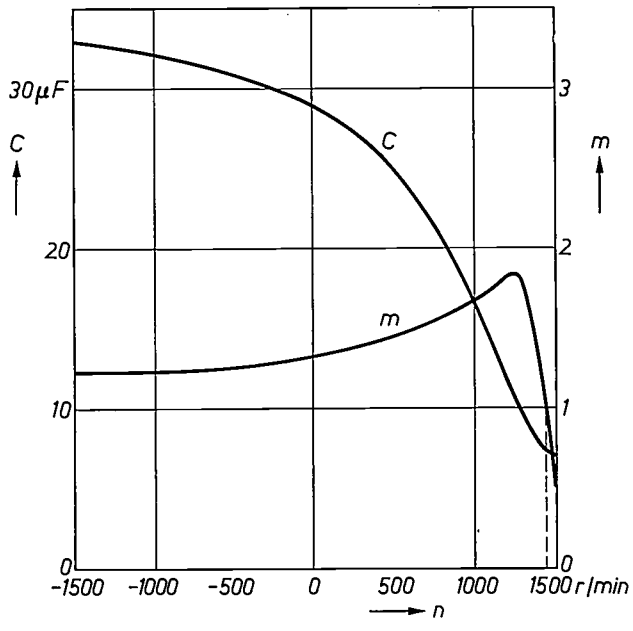


Fig. 7. Calculation on the basis of fig. 6 of the capacitance C of the capacitor and of the turns ratio m of the autotransformer required in order to balance a 370 W motor; both are plotted as a function of the speed n. For $n \approx 1450$ rev/min the value of m is 1 and the autotransformer is therefore unnecessary.

number of turns, it must be 45° in both phases. This phase angle is best approximated at the nominal speed and at zero or negative speed. *Fig. 5*, which gives a vector representation of the current in a three-phase motor for various speeds, shows that the phase angle is identical in these two regions. For starting a non-controlled motor it is the practice to use a starting capacitor that balances the motor at zero speed. Once the motor has reached its nominal speed, a switch is made to another capacitor of smaller value. This balances the motor approximately at the nominal speed.

If instead of a capacitor we use a combination of a capacitor with a resistor or inductor for the auxiliary impedance, or we give the main and auxiliary windings different numbers of turns, then in principle we can obtain electrical balance for any given speed, but only for one speed if the elements have fixed values.

This is illustrated by the circuit diagram and the associated vector diagram of a capacitor motor (*fig. 6*). A capacitor and an autotransformer are used for achieving electrical balance. The vector diagram is constructed starting from the mains vo ltage V_m, which lies between the terminal points R and S of the motor. Assuming that the motor is balanced, the voltages with respect to the star point are of equal magnitude ($V_{RM} = V_{SM} = V_{TM}$) and so also are the phase currents ($I_R = I_S = I_T$). The angle between current and voltage is read from the current curve of the machine (fig. 5), which has previously been recorded. Making it a condition that I_C must be perpendicular to the capacitor voltage V_C, we can then construct V_C and the transformer voltage V_{QS}. The capacitor reactance $X_C = V_C/I_C$ and the turns ratio $m = V_{QS}/V_{RS}$ can then be determined.

This turns ratio m and the capacitance C of the capacitor were determined for a 370 W motor as a function of the motor speed (*fig. 7*). At $n \approx 1450$ rev/min $m = 1$, so that the autotransformer is not necessary. Symmetry at all other speeds in the control range can only be obtained with $m \neq 1$.

Optimum value of the capacitor

We now come to the first of the questions with which this article is concerned: what is the optimum value of the capacitor for use in the four-quadrant motor? To answer this question we used a computer to calculate the torque at various speeds and capacitor reactances (*fig. 8*). We assumed a constant rotor resistance and neglected the skin effect and saturation of the iron. The torque was normalized to the pull-out (maximum) torque T_{max} which the motor reaches when it is connected to a three-phase supply. The characteristic of the machine in this case is shown by a dashed line in fig. 8. (In determining T_{max} we neglected the stator resistance and assumed a leakage coefficient $\sigma = 0.15$.)

If we take the capacitor reactance X_C as equal in magnitude to, for example, 25% of the reactance X of the unloaded machine, then the torque-speed characteristic of the single-phase fed machine in fig. 8*a* lies below that of the symmetrical machine only for $n > 0.8\, n_0$. In this range the motor with this capacitor cannot in any case be used because of the torque pulsations that occur and the noise they cause. At $X_C = 0.3\, X$ the rotating field of opposite sign largely disappears for $n = 0.7\, n_0$. This advantage is offset by the decrease of the torque at $n = -0.8\, n_0$. At $X_C = 0.35\, X$ the torque even becomes negative in the range $-0.9\, n_0 < n < -0.75\, n_0$.

Fig. 8*b* gives a plot of the torque with the capacitor reactance for positive and negative speeds and for zero speed. In seeking to establish a torque that is constant and as large as possible in the whole speed-control range the optimum choice is found to be $X_C = 0.2$ to $0.3\, X$; this interval represents approximate electrical balance for zero and negative speeds.

Torque pulsations

So far we have been considering the static torque of induction motors with an auxiliary phase. However, as we have seen, 100-Hz pulsations occur in the torque as soon as the speed deviates from the value at which the

motor is electrically balanced. These pulsations arise because the stator-field components rotating in the sense opposite to the rotor-field components generate a torque alternating at 100 Hz, which is superimposed on the static torque.

In *fig. 9* the calculated amplitude of this alternating torque is plotted against n (again normalized to the pull-out torque of the symmetrical machine). At zero speed there is of course no alternating torque, whether the machine is symmetrical for $n = 0$ or not. At $X_C = 0.3\,X$ and $n = 0.7\,n_0$ (fig. 9*b*) the alternating torque has a minimum, since the field rotating in the opposite sense has almost vanished at this point (see also fig. 8*a*). In the interval $0.85\,n_0 < n < n_0$ the alternating torque is greater than the static torque; the motor runs very roughly here.

A comparison of fig. 9*a* and *b* shows that the ratio of the alternating to the static torque with increasing capacitor reactance becomes worse in the braking range and better in the motor range. The speed at which balance is optimum shifts to higher values. At very high reactances (fig. 9*c*) the machine approximates to the behaviour of a single-phase machine. The alternating torque is then of the same order of magnitude as the static torque.

Here again, the value $X_C = 0.25\,X$ (fig. 9*a*) is an optimum choice and leads to minimum torque pulsations in the whole of the control range.

Dips in the torque-speed characteristic caused by the third harmonic in the field distribution

Commercially available capacitor motors are usually two-phase machines. The advantage is that exact balance can be reached even from a phase angle of $\pi/4$ with a resistor and a capacitor in series.

The torque-speed characteristics of a conventional 370 W motor with auxiliary phase and starting capacitor are shown in *fig. 10*. The motor is balanced for $n = 0$. As can be seen, at $n = \pm \frac{1}{3}\,n_0$ there is a conspicuous dip in the characteristics in the motor and braking ranges. This is due to the auxiliary winding of the two-phase machine. It is this winding that is mainly responsible for a non-sinusoidal distribution of the field in the air gap; in addition to a component corresponding to the number of poles in the machine (the fundamental wave) the field also has a third-harmonic component corresponding to three times the number of poles. (In three-phase machines the third-harmonic components produced by each of the phases cancel one another out.) The third-harmonic field component can be treated as a standing wave and resolved into two waves rotating in opposite senses, one at a velocity of $\frac{1}{3}\,n_0$, the other at $-\frac{1}{3}\,n_0$. Both these waves produce torques that pass through zero on the horizontal axis at $n = \pm \frac{1}{3}\,n_0$; these torques are added to the torque of the fundamental wave. The torque produced by the third-harmonic wave rotating in the same sense is a

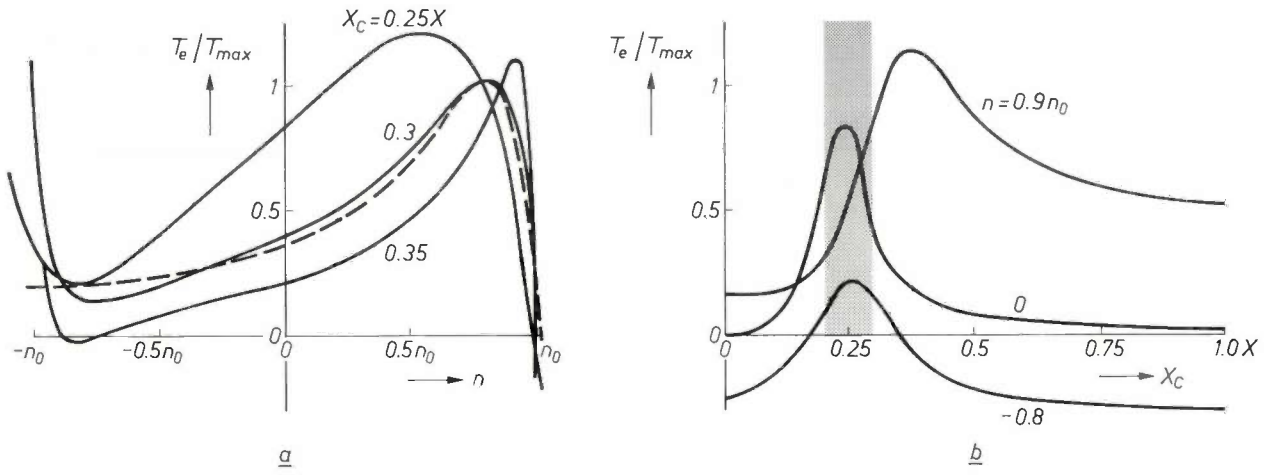


Fig. 8. *a*) Calculated torque-speed characteristics of a capacitor motor for various reactances X_C of the capacitor. The torque T_e is expressed in terms of the pull-out torque T_{max} with a symmetrical supply. The dashed line indicates the torque-speed characteristic of the symmetrically fed motor. *b*) Torque as a function of capacitor reactance for three speeds. In the shaded region the torque is also positive at negative speeds and is reasonably high both at positive and zero speeds; the optimum choice of capacitor reactance lies in this region.

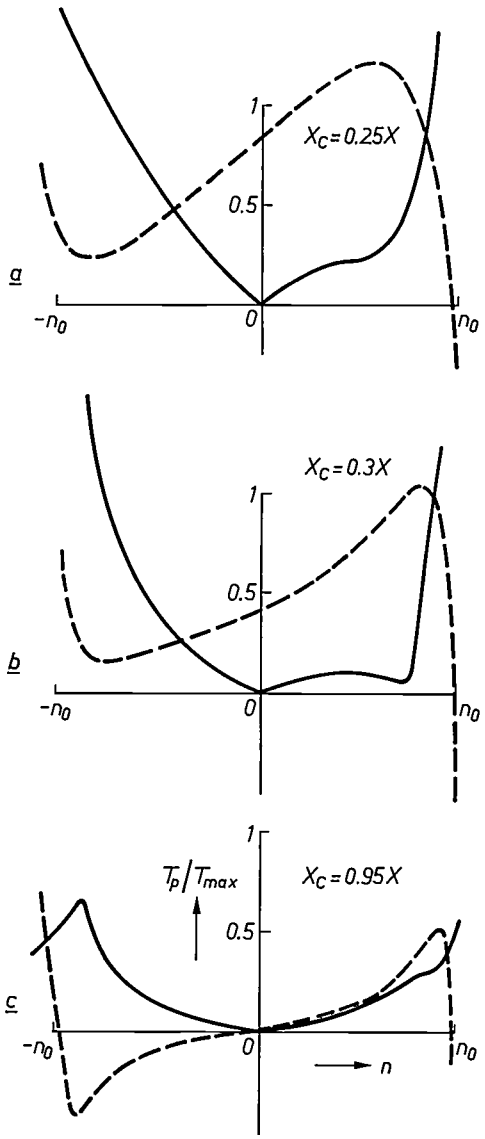


Fig. 9. Calculated amplitude T_p of the 100-Hz torque fluctuations of a capacitor motor as a function of the speed (T_p is normalized at the pull-out torque T_{max} for symmetrical supply). The static torque is indicated by a dashed line. *a*) $X_C = 0.25\,X$. *b*) $X_C = 0.3\,X$. At $n = 0.7\,n_0$ the torque fluctuations are largely eliminated because the motor is practically symmetrical at that speed. *c*) $X_C = 0.95\,X$. At this high reactance of the capacitor the machine approximates to the behaviour of a single-phase induction motor.

motor torque in the speed range $-n_0 < n < \frac{1}{3}n_0$ where it increases the total torque, but it is a generator torque in the range $\frac{1}{3}n_0 < n$, where it reduces the total torque.

The torque reduction caused by the auxiliary winding causes no trouble in the applications for which such a motor is designed, since the rated torque that the motor reaches from the start is always greater than the torque at the dip and because there is no speed control. If a drive does have speed control, however, the effect of these third harmonics in the field will be to make its dynamic behaviour strongly dependent on the speed, and in the rising part of the torque-speed characteristic it will have a tendency to 'hunt'.

If there is no alternative but to use two-phase machines, it is as well to find a motor in which the coil width of the main and auxiliary windings has been shortened to two-thirds of the pole pitch. In this way the third harmonic of the field can be cancelled out exactly in the air gap. In practice, however, it is usually advisable not to use two-phase machines (or asymmetrical three-phase machines) but instead commercially available three-phase machines in star connec-

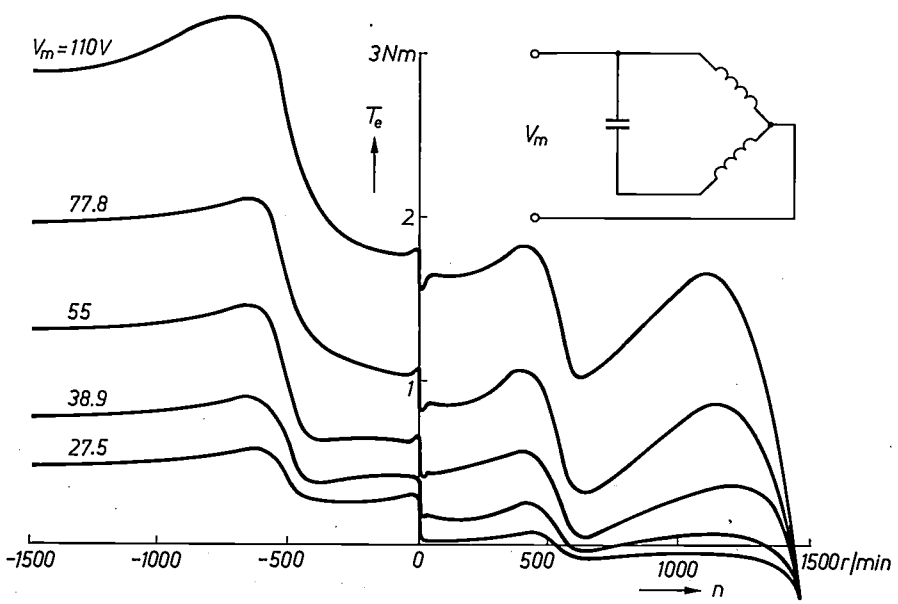


Fig. 10. Torque-speed characteristics of a two-phase induction motor with starting capacitor. Marked dips at $n = \pm\frac{1}{3}n_0$ are caused by the third harmonic in the field distribution of the auxiliary winding. These irregularities interfere with the operation of a speed-controlled motor.

tion. In the first place this avoids the third field harmonic, and in the second place there is the advantage that these 220/380 V motors connected to a 220 V mains are then conservatively rated. They are less likely to be used for very small conduction angles α, where the higher-harmonic content is unfavourably large, and they are better able to withstand the considerable heat development that is unavoidable at low speeds (see below).

Stability of the system at small conduction angles

We shall now take a closer look at the stability of the speed-control system. We are not concerned here with the chance of instability that exists in every feedback system, but with a special form of spontaneous oscillation observed at small conduction angles of the control thyristors used with capacitor motors.

If the conduction angle α is smaller than a critical value, the measured torque-speed characteristic deviates from the normal curve (*fig. 11*). Above a particular speed ($|n| > 1000$ rev/min) additional braking torques then occur in both motor and braking ranges. These additional torques are not due to currents and rotating fields originating from the mains but to a spurious oscillation.

In describing this oscillation we shall proceed from the limiting case in which the motor is entirely separated from the mains by the thyristors ($\alpha = 0$). In theory (disregarding the losses) this is the case when the motor is run on zero load. This does not imply that the speed is necessarily equal to the synchronous speed; for even in the absence of a load a speed-controlled motor keeps to a constant speed n as long as the system is electrically stable. In practice the conduction angle in the absence of a load is not quite zero, but it can nevertheless be postulated that the internal resistance R_i of the mains is infinitely high for the motor.

Fig. 12a shows the thyristor-controlled three-phase motor used in the stability investigation. During the investigation we opened the control loop and ran the capacitor motor up to a particular speed by means of a drive motor M_a. The torque was controlled by changing the conduction angle of the thyristors, which also had the effect of changing the internal resistance of the mains as seen by the motor.

Fig. 12*b* shows the equivalent circuit used for explaining the origin of the oscillations. In addition to the inductances and resistances of the machine the equivalent circuit contains only the capacitance C and possibly a resistance R_p connected in parallel with it. The mains voltage V_m and the mains frequency are not included, since we assume that the machine is completely isolated from the mains.

The stator of the machine can only excite a standing wave in the air gap, since only two terminals carry current. This standing wave can be resolved into a component travelling with the rotor and a component travelling in the opposite sense. In the equivalent circuit this corresponds to a series arrangement of two impedances. The voltage across the upper impedance (fig. 12*b*) is equal to the voltage generated at the motor terminals by the stator and rotor fields rotating in the same direction, while the voltage across the lower impedance is due to the opposite fields. These impedances comprise the transformed rotor resistance $R_r'/(1 \pm n/n_0)$ with a minus sign in the denominator for the stator component field that has the same sense of rotation as the rotor, and a plus sign for the field of the opposite sense. It should be noted that in this context n_0 is no longer connected with the mains frequency but with the frequency of the a.c. currents arising spontaneously in the stator.

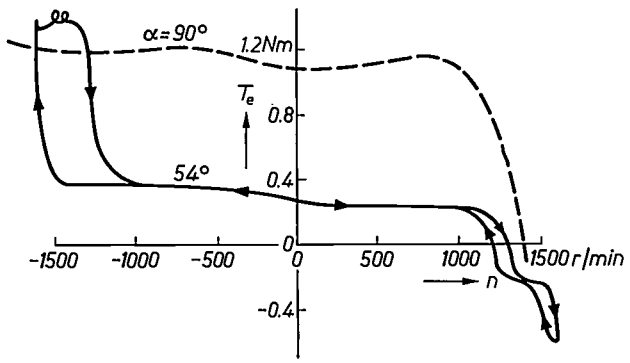


Fig. 11. Torque-speed characteristics measured on a thyristor-controlled capacitor motor with a conduction angle $\alpha = 54°$ (solid curve) and $\alpha = 90°$ (dashed). The solid curve shows irregularities at speeds $|n| > 1000$ rev/min, which are the result of electrical oscillations at frequencies at which the circuit formed by the capacitor and the inductance of the motor resonates. At larger conduction angles the damping from the mains is greater and the oscillations do not occur (dashed curve).

The resistance $R_r'/(1 - n/n_0)$ in fig. 12*b* assumes negative values as soon as the rotor starts to rotate faster than the stator field, i.e. when $n > n_0$. Because of this negative resistance the positive resistances R_s, R_p and $R_r'/(1 + n/n_0)$ are compensated, which removes the damping of the system.

There are two possible resonant frequencies, which occur at $n \approx n_0$ and at $n \gg n_0$. If $n \approx n_0$, we can neglect branch *1* in fig. 12*b* compared with branch *2*, and substitute a short-circuit for branch *3*. The capacitance C then resonates with the inductance L_s of the unloaded motor, and the resonant frequency $\omega_{01}/2\pi$ is given approximately by $\omega_{01} \approx 1/\sqrt{L_s C}$. If $n \gg n_0$, then the branches *2* and *4* can be neglected compared with *1* and *3*; the capacitance then resonates with the

distributed inductances, and the resonant frequency is given by

$$\omega_{02} \approx \frac{1}{\sqrt{2(L_{s\sigma} + L_{r\sigma}')C}}.$$

It can be seen from fig. 11 that the spontaneous oscillations only occur above a critical speed. As in the case of a linear oscillator, they are excited by the thermal noise that is always present in the system. The magnetic saturation of the iron is the nonlinearity that limits the amplitude. Depending on the speed and size of the machine, this limitation of voltage and current amplitude may, however, be far above the rated values of the machine. The currents arising could well burn out the windings, since considerable power is dissipated in the resistances of the stator. In the circuit used for the experiment this power is delivered by the drive motor. If the torque of this motor is also high at low speeds, and if it has a high moment of inertia, the sudden occurrence of the oscillation can even shear off the motor shaft.

Prevention of the spontaneous oscillations

The excitation of spontaneous oscillations can be prevented by inserting damping resistances in the stator circuit. This is why there is a damping resistor in fig. 12 in parallel with the capacitor. The oscillations can also be suppressed, however, by increasing the resistance R_s of the stator circuit by connecting resistances in series with it.

From an equivalent circuit corresponding to that in fig. 12*b* an equation can be derived for the critical value of the resistance R_s:

$$R_{s,cr} \approx \tfrac{1}{2}(L_s/C)^{\frac{1}{2}}[1 + 3\sigma - 2\{2\sigma(1+\sigma)\}^{\frac{1}{2}}]^{\frac{1}{2}},$$

where σ is the leakage coefficient and is given by:

$$\sigma = (L_{s\sigma} + L_{r\sigma}')/L_s.$$

If the resistance of a phase is greater than $R_{s,cr}$, spontaneous oscillation cannot occur.

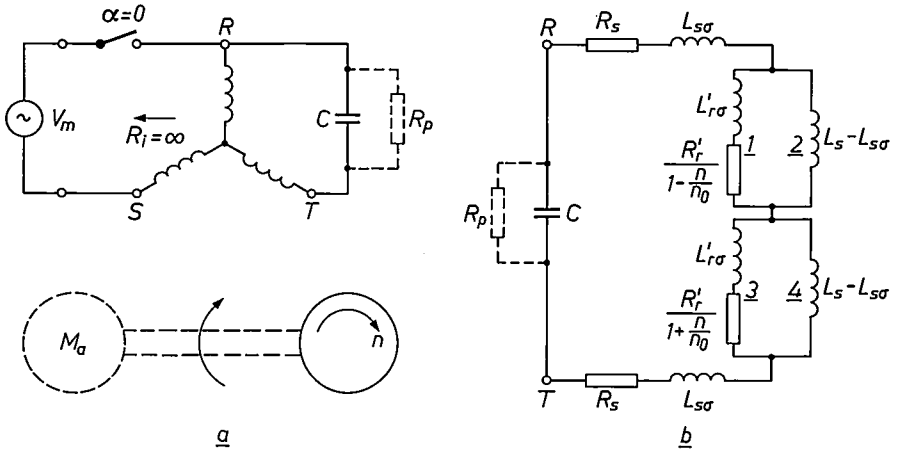


Fig. 12. *a*) Capacitor-motor circuit for studying instabilities. A separate drive motor M_a gives the capacitor motor a speed n. At $\alpha = 0$ the internal resistance R_i of the mains as seen from the motor is infinitely high. *b*) Equivalent circuit of the capacitor motor with capacitor C (and possibly a parallel resistance R_p) when α is put at 0. The parallel circuit *1,2* represents the effect of the stator-field component rotating with the rotor field, the parallel circuit *3,4* represents that of the other component rotating in the opposite sense. R_s d.c. resistance and $L_{s\sigma}$ leakage inductance of a motor phase. L_s inductance of a phase in no-load operation of the motor. R_r' d.c. resistance and $L_{r\sigma}'$ leakage inductance of the rotor transformed to the stator circuit. The resistance $R_r'/(1 - n/n_0)$ can assume negative values and reduce the damping of the circuit sufficiently to enable spontaneous oscillations to appear.

In some cases, however, it is preferable to connect a resistance in parallel with the input rather than a resistance $R_C \approx 2\,R_{s,cr}$ in series with the capacitor. The advantage is that the symmetry of the rotating field under full load ($\alpha = 180°$) is not then disturbed, and moreover the losses in the parallel resistance are smaller.

No damping resistance required at low powers

If the nominal power P_N of a capacitor motor is lower than a critical value $P_{N,cr}$, no damping resistance is necessary to prevent spontaneous oscillation. The smaller the motor the greater the d.c. resistance in relation to the reactance; below the limit $P_{N,cr}$ the resistance of the stator winding exceeds the value $R_{s,cr}$, and the motor is therefore sufficiently damped.

The important factor is not the absolute magnitude of R_s but the ratio X/R_s. *Fig. 13* shows how this ratio increases with the power of the motor; the graph is based on a study of catalogue data and on measure-

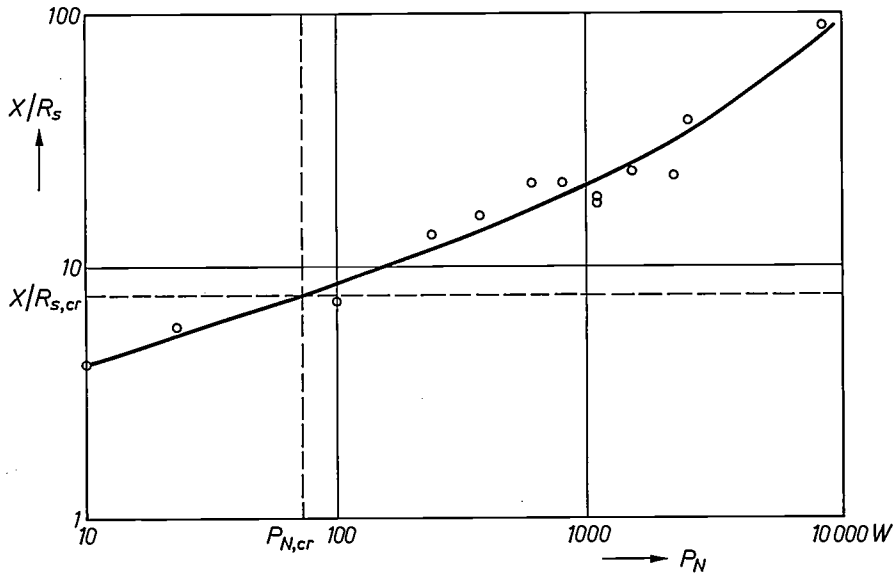


Fig. 13. The ratio of the stator reactance X (measured for unloaded motor) to the stator resistance R_s in motors with a nominal power P_N between 10 W and 10 kW. Below a power $P_{N,cr}$ of about 70 W this ratio remains below the critical value $X/R_{s,cr} = 7.6$ (for $\sigma = 0.1$) and no spontaneous oscillation can occur. Above this power a damping resistance is required.

ments performed on motors available. It can be seen that at lower power ratings X/R_s increases approximately as $P_N^{\frac{1}{4}}$, which is easy to prove from analysis.

From the expression for $R_{s,cr}$ given above, an expression can be derived for the critical ratio $X/R_{s,cr}$; if the ratio X/R_s is less than $X/R_{s,cr}$ the motor is stable. Using the relations $X = \omega_0 L_s$ and $X_C = 1/\omega_0 C$ we find:

$$X/R_{s,cr} \approx 2(X/X_C)^{\frac{1}{2}} [1 + 3\sigma - 2\{2\sigma(1+\sigma)\}^{\frac{1}{2}}]^{-\frac{1}{2}}.$$

Assuming a capacitor reactance $X_C = 0.2\,X$ (see fig. 8) and putting the leakage coefficient σ at 0.1 (in the machines of interest here its magnitude lies between 0.08 and 0.15) we arrive at the following value:

$$X/R_{s,cr} = 7.6.$$

This value is indicated in fig. 13 by a dashed line. The critical power $P_{N,cr}$ lies at the point where this line cuts the characteristic, and can be seen to be about 70 W. Above this value the ratio quickly deteriorates; at 8 kW, for example, the ratio X/R_s is already ten times too large.

If the power P_N is greater than $P_{N,cr}$, spontaneous oscillation can only occur if the speed is in the range determined by R_s, R_r, L_s and C. At powers from 70 to 150 W this range is very small, so that in many applications there is no need for special damping precautions.

Spontaneous oscillations during speed control

So far we have considered the spontaneous oscillations in the case where the speed is set from outside by a drive motor. When the speed control is operative, however, the speed must adjust itself to the set value. The dynamic behaviour of the system then comes into play; the conduction angle α will be between 0° and 180°. If there is now no damping resistance to suppress the oscillations, complex interactions will arise between the torque caused by the oscillations, which is always a braking torque, and the torque produced by the 50-Hz mains currents. This causes extra heating in the motor, as well as noise and possible speed fluctuations, and therefore the damping resistance cannot be omitted.

Heating in the motor during phase control

In a phase-controlled induction motor the considerable losses that occur at high slip can be a real problem. In a wound-rotor machine the heat generated can easily be dissipated through the starting resistor, which is connected to the rotor externally, but in squirrel-cage motors the heat has to be removed directly from the surfaces of the machine. This means that in continuous operation the machine may only be loaded with a fraction of the nominal torque. In intermittent operation the motor may be loaded briefly with the nominal torque or even a higher torque, but then it must be switched off and left to cool.

We have performed measurements to determine the torque a 370 W motor can deliver continuously at any speed when the temperature of the windings is not allowed to exceed a specified value. This temperature was kept constant at the maximum permissible value [3] by means of the thyristor control and the delivered torque was measured at different speeds. The fins of the motor housing were cooled either by the ventilator fan on the motor shaft or by an independent fan running at a fixed speed.

[3] This was set at 85 °C above ambient temperature at the hottest place, corresponding to insulation class E in the regulations of the German Electrical Engineers Association (VDE, *Verband deutscher Elektrotechniker*).

The torque-speed characteristics thus measured are shown in *fig. 14*; the lower curve relates to cooling with the built-in fan, the upper curve to external ventilation. Fig. 14*a* shows the results for supply from three-phase mains and control in all three phases; it can be seen that at zero speed only 26% of the nominal torque is continuously available even with external ventilation. With a single-phase supply and control (fig. 14*b*) the percentage is even smaller: only 21%. Furthermore the permitted torque then hardly increases at all with rising speed. This is because the motor was balanced with a transformer and capacitor for zero speed and at higher speeds the rotary field is therefore asymmetrical.

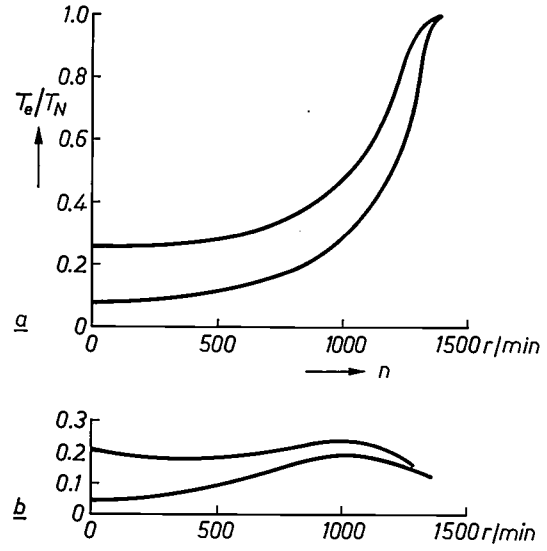


Fig. 14. Permitted torque T_e as a function of the speed n for a continuously loaded thyristor-controlled motor; when the torque is exceeded the motor becomes overheated. T_e is normalized to the nominal torque T_N. Upper curve: constant air cooling provided by a separate ventilator fan; lower curve: cooling by ventilator fan on the motor shaft. *a*) Symmetrical supply and control in three phases. At low speeds n the losses in the motor are high and the permitted torque T_e is much lower than T_N. At high speeds it approaches T_N. *b*) Capacitor motor. This was balanced for $n = 0$, and the asymmetry of the field at high speeds causes additional losses. Here again the permitted torque T_e remains far below T_N.

A capacitor motor in intermittent use, if it is to be loaded with the nominal torque at low speeds, must be switched off for five times as long as it is switched on (e.g. 10 seconds on and 50 seconds off, depending on the thermal time constants). Generally, it is even more important than with a three-phase controlled motor to ensure that ventilation and cooling for a speed-controlled capacitor motor are adequate for the operating conditions.

Summary. The speed of induction motors with a capacitor can be regulated by phase control of the stator voltage. This is usually done by means of thyristors. Below a power of about 800 W these capacitor motors can readily be used either in one or in four quadrants, with the advantages that only a few electronic components are required and a single-phase mains supply can be used. A capacitor motor is particularly suitable for drives where smooth running is not a first requirement and the motor only has to be switched on for a short time. In the design of such drives it is necessary to make the rotating field symmetrical, to avoid the third harmonic in the field distribution and to ensure stability at low loads. Three-phase motors can best be used in star connection and balanced for zero speed. At powers above 100 W a damping resistance is required. To ensure smooth running and low losses it is advisable to set the upper limit to the speed range at about 80% of the unloaded speed.

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.

G. Armand (Service de Physique Atomique, Gif-sur-Yvette) **& J. B. Theeten:** Surface phonons in $C(2\times2)$ adsorbed layers on Ni(001): a criterion for distinguishing between reconstructed and non-reconstructed layers.
Solid State Comm. **13**, 563-568, 1973 (No. 5). *L*

C. Belouet: Croissance en solution aqueuse, I. Considérations générales, II. Croissance de KH_2PO_4 par la méthode de descente en température.
Acta Electronica **16**, 339-353, 1973 (No. 4). *L*

N. Bloembergen: The influence of electron plasma formation on superbroadening in light filaments.
Optics Comm. **8**, 285-288, 1973 (No. 4). *E*

G. M. Blom & W. K. Zwicker: The growth of GaP single crystals by liquid encapsulated Czochralski pulling.
Acta Electronica **16**, 315-322, 1973 (No. 4). *N*

A. H. Boonstra & R. M. A. Sidler: Partial substitution of oxygen in the surface layer of vapor-deposited lead monoxide crystallites by chemisorption of hydrogen chloride.
J. Electrochem. Soc. **120**, 1078-1083, 1973 (No. 8). *E*

M. Bouckaert, A. Pirotte & M. Snelling: Improvements to Earley's context-free parser.
Lecture Notes in Computer Science **1**, 104-112, 1973 (Springer, Berlin). *B*

J. C. Brice: Controlling heat transport during crystal pulling.
Acta Electronica **16**, 291-301, 1973 (No. 4). *M*

J.-J. Brissot: A history of crystals.
Acta Electronica **16**, 285-290, 1973 (No. 4). *(Also in French, pp. 279-284.)* *L*

A. Broese van Groenou: Nachwirkungsmechanismen in Ferriten.
Appl. Phys. **2**, 47-58, 1973 (No. 2). *E*

E. Bruninx: The accurate determination of major components in Ga_xSe_y by means of instrumental neutron activation.
Anal. Chim. Acta **67**, 17-28, 1973 (No. 1). *E*

T. M. Bruton: Study of the liquidus in the system Bi_2O_3-TiO_2.
J. Solid State Chem. **9**, 173-175, 1974 (No. 2). *M*

K. H. J. Buschow: Magnetic properties of CsCl-type rare earth-magnesium compounds.
J. less-common Met. **33**, 239-244, 1973 (No. 2). *E*

K. H. J. Buschow: Magnetic anisotropy of some rare earth-cobalt compounds (R_2Co_7).
J. less-common Met. **33**, 311-312, 1973 (No. 2). *E*

K. H. J. Buschow & F. J. A. den Broeder: The cobalt-rich regions of the samarium-cobalt and gadolinium-cobalt phase diagrams.
J. less-common Met. **33**, 191-201, 1973 (No. 2). *E*

P. A. Devijver: Relationships between statistical risks and the least-mean-square-error design criterion in pattern recognition.
Proc. 1st Int. Joint Conf. on Pattern Recognition, Washington 1973, pp. 139-148. *B*

J. Dieleman, A. W. Witmer, J. C. M. A. Ponsioen & C. P. T. M. Damen: Rapid and inexpensive sampling technique for emission spectroscopic analysis of thin films.
Appl. Spectr. **27**, 387-388, 1973 (No. 5). *E*

G. Dittmer, A. Klopfer, D. S. Ross & J. Schröder: Transport reactions in the tungsten fluorine system.
J. Chem. Soc., chem. Comm., 1973, 846-847 (No. 22). *A*

E. Dormann (Technische Hochschule Darmstadt) **& K. H. J. Buschow:** The hyperfine fields in ferromagnetically ordered cubic Laves phase compounds of gadolinium with non-magnetic metals.
Phys. Stat. sol. (b) **59**, 411-418, 1973 (No. 2). *E*

D. den Engelsen: Ellipsometry of fluid interfaces and membrane-like systems.
Chemie-Ing.-Technik **45**, 1107-1109, 1973 (No. 18). *E*

D. den Engelsen: Monolayers and multilayers of arachidic acid with rhodamine 6G.
J. Colloid & Interface Sci. **45**, 1-10, 1973 (No. 1). *E*

C. T. Foxon: Molecular beam epitaxy.
Acta Electronica **16**, 323-329, 1973 (No. 4). *M*

K. L. Fuller: Solid-state radar.
Electronics & Power **20**, 100-101, 1974 (21 Feb.). *M*

Z. van Gelder & M. M. M. P. Mattheij: Principles and techniques in multicolor dc gas discharge displays.
Proc. IEEE **61**, 1019-1024, 1973 (No. 7). *E*

C. J. Gerritsma & J. H. J. Lorteye: A hybrid liquid-crystal display with a small number of interconnections.
Proc. IEEE **61**, 829-832, 1973 (No. 7). *E*

G. G. P. van Gorkom: Doubly excited Cr^{3+} pairs in $ZnGa_2O_4$.
Phys. Rev. B **8**, 1827-1834, 1973 (No. 5). *E*

J. Graf: Les multiplicateurs d'électrons à microcanaux.
Electronique & Microél. ind. No. 176, 33-38, 1973. *L*

S. H. Hagen, A. W. C. van Kemenade & J. A. W. van der Does de Bye: Donor-acceptor pair spectra in 6H and 4H SiC doped with nitrogen and aluminium.
J. Luminescence **8**, 18-31, 1973 (No. 1). *E*

J. 't Hart & A. Cohen (Institute for Perception Research, Eindhoven): Intonation by rule: a perceptual quest.
J. Phonetics **1**, 309-327, 1973.

E. E. Havinga, K. H. J. Buschow & H. J. van Daal: The ambivalence of Yb in $YbAl_2$ and $YbAl_3$.
Solid State Comm. **13**, 621-627, 1973 (No. 5). *E*

H. van der Heide: Dimensional considerations concerning lifting forces of magnetically levitated trains.
Philips Res. Repts. **29**, 152-154, 1974 (No. 2). *E*

J. C. M. Henning & H. van den Boom: ESR investigations of nearest-neighbor Cr^{3+}-Cr^{3+} interactions in Cr-doped spinel $MgAl_2O_4$.
Phys. Rev. B **8**, 2255-2262, 1973 (No. 5). *E*

W. K. Hofker (Philips Research Labs., Amsterdam Division), **H. W. Werner, D. P. Oosthoek** (Philips Res. Labs. Amsterdam) **& H. A. M. de Grefte:** Influence of annealing on the concentration profiles of boron implantations in silicon.
Appl. Phys. **2**, 265-278, 1973 (No. 5). *E*

E. P. Honig, J. H. Th. Hengst & D. den Engelsen: Langmuir-Blodgett deposition ratios.
J. Colloid & Interface Sci. **45**, 92-102, 1973 (No. 1). *E*

H. Kalis: Phasenregelkreise in der Prozeß-Automatisierungstechnik.
Elektronik **22**, 379-382 & 390, 1973 (No. 11). *H*

D. Kasperkovitz: A dynamic delay line with a bipolar one-transistor cell.
IEEE J. **SC-8**, 251-259, 1973 (No. 4). *E*

K.-G. Knauff, G.-A. Lens & A. Wiericks: Steuerelektronik für hochpräzise automatische Titrationsanalyse.
Int. elektron. Rdsch. **27**, 273-275, 1973 (No. 12). *A*

J. E. Knowles: An apparatus to determine magneto-crystalline anisotropy as a function of frequency in the range 2 Hz to 50 kHz.
J. Physics E **7**, 91-94, 1974 (No. 2). *M*

J. E. Knowles: Magnetic after-effects in ferrites substituted with titanium or tin.
Philips Res. Repts. **29**, 93-118, 1974 (No. 2). *M*

A. J. R. de Kock: Microdefects in dislocation-free silicon and germanium crystals.
Acta Electronica **16**, 303-313, 1973 (No. 4). *E*

J.-P. Krumme & P. Hansen: New magneto-optic memory concept based on compensation wall domains.
Appl. Phys. Letters **23**, 576-578, 1973 (No. 10). *H*

P. K. Larsen & R. Metselaar: Non-ohmic currents in inhomogeneous polycrystalline yttrium iron garnet.
Mat. Res. Bull. **8**, 883-892, 1973 (No. 8). *E*

P. K. Larsen & R. Metselaar: Electric and dielectric properties of polycrystalline yttrium iron garnet: space-charge-limited currents in an inhomogeneous solid.
Phys. Rev. B **8**, 2016-2025, 1973 (No. 5). *E*

A. R. Miedema: The electronic heat capacity of transition metal solid solutions: an alternative to the rigid band model I.
J. Physics F **3**, 1803-1818, 1973 (No. 10). *E*

K. Mouthaan: Transmissie over optische kabels voor de lange en middellange afstand.
T. Ned. Elektronica- en Radiogen. **38**, 113-122, 1973 (No. 5). *E*

B. J. Mulder: Optical properties and energy band scheme of cuprous sulphides with ordered and disordered copper ions.
Phys. Stat. sol. (a) **18**, 633-638, 1973 (No. 2). *E*

P. A. Naastepad (Metallurgical Laboratory, Philips P. M. F. Division, Eindhoven), **F. J. A. den Broeder & R. J. Klein Wassink:** Technology of $SmCo_5$ magnets.
Powder Metall. Int. **5**, 61-65, 1973 (No. 2). *E*

A. K. Niessen & A. den Ouden: Conoscopic observations on some smectic liquid-crystalline materials.
Philips Res. Repts. **29**, 119-138, 1974 (No. 2). *E*

K. J. van Oostrum, A. Leenhouts & A. Jore: A new scanning microdiffraction technique.
Appl. Phys. Letters **23**, 283-284, 1973 (No. 5). *E*

A. Oppelt (Technische Hochschule Darmstadt) **& K. H. J. Buschow:** Y hyperfine fields in YFe_2, YFe_3 and Y_2Fe_{17}.
J. Physics F **3**, L 212-215, 1973 (No. 10). *E*

R. Polaert & J. Rodière: Internal investigation of microchannel plates by scanning electron microscopy.
Rev. sci. Instr. **44**, 1531-1536, 1973 (No. 10). *L*

J. M. Robertson, S. Wittekoek, Th. J. A. Popma & P. F. Bongers: Preparation and optical properties of single crystal thin films of bismuth substituted iron garnets for magneto-optic applications.
Appl. Phys. **2**, 219-228, 1973 (No. 5). *E*

F. Rondelez: Contribution à l'étude des effets de champ dans les cristaux liquides nématiques et cholestériques.
Thesis, Paris-Sud 1973. (Philips Res. Repts. Suppl. 1974, No. 2.) *L*

U. Rothgordt: The influence of the contact impedance between base paper and back electrode on the electrostatic recording process.
Philips Res. Repts. **29**, 139-151, 1974 (No. 2). *H*

B. Schiek: Stabilization factor of a cavity-controlled microwave oscillator with several output ports.
Arch. Elektronik & Übertr.technik (AEÜ) **27**, 490-491, 1973 (No. 11). *H*

A. J. Smets: The fine sun sensor of the astronomical Netherlands satellite.
Industries atom. & spat. **17**, No. 3, 77-82, 1973. *E*

J. H. Statius Muller: Programmering van minicomputers.
Informatie **15**, 458-463, 1973 (No. 9). *E*

T. J. B. Swanenburg: Negative conductance of an interdigital electrode structure on a semiconductor surface.
IEEE Trans. **ED-20**, 630-637, 1973 (No. 7). *E*

T. J. B. Swanenburg & J. Wolter: Transmission of high-frequency phonons through a solid-liquid-helium interface.
Phys. Rev. Letters **31**, 693-696, 1973 (No. 11). *E*

A. Thayse: Applications of discrete functions, Part II. Transient analysis of asynchronous switching networks.
Philips Res. Repts. **29**, 155-192, 1974 (No. 2). *B*

H. Uhlemann: Die Eigenschaften von Drahtgewebestrukturen als Flüssigkeitsverteiler in Dünnschichtverdampfern.
Thesis, Eindhoven 1974. (Philips Res. Repts. Suppl. 1974, No. 1.) *E*

A. A. van der Veeke: Wide-range linear or exponential frequency control of an astable multivibrator.
Electronic Engng. **45**, Nov. 1973, 13 (No. 549). *E*

J. van der Veen, W. H. de Jeu, M. W. M. Wanninkhof (Philips Elcoma Division, Eindhoven) **& C. A. M. Tienhoven** (Philips Elcoma Division, Eindhoven): Transition entropies and mesomorphic behavior of para-disubstituted azoxybenzenes.
J. phys. Chem. **77**, 2153-2155, 1973 (No. 17). *E*

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven), **J. L. Sommerdijk & J. G. Verriet** (Philips Lighting Division, Eindhoven): Cerium and terbium luminescence in $LaMgAl_{11}O_{19}$.
J. Luminescence **6**, 425-431, 1973 (No. 5). *E*

A. T. Vink, R. L. A. van der Heyden & J. A. W. van der Does de Bye: The dielectric constant of GaP from a refined analysis of donor-acceptor pair luminescence, and the deviation of the pair energy from the Coulomb law.
J. Luminescence **8**, 105-125, 1973 (No. 2). *E*

L. Vriens: Energy balance in low-pressure gas discharges.
J. appl. Phys. **44**, 3980-3989, 1973 (No. 9). *E*

J. P. Woerdman: A new interpretation of the strain-splitting of bound-exciton lines in CdS.
Solid State Comm. **13**, 949-951, 1973 (No. 7). *E*

W. K. Zwicker & S. K. Kurtz: The growth of silver and copper single crystals on silicon and the selective removal of silicon by electrochemical displacement.
Acta Electronica **16**, 331-338, 1973 (No. 4). *N*

Contents of Philips Telecommunication Review **32**, No. 2, 1974:

Contents of Mullard Technical Communications **13**, No. 123, 1974:

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 8

Materials research for permanent magnets

H. Zijlstra

Permanent-magnet material is used very widely in technical products. The total world turnover in this material in 1973 was estimated at more than five hundred million guilders. The demand for permanent magnets, particularly for small yet powerful types, is steadily increasing. They have innumerable applications: they are used in small electric motors, dynamos, relays, in all kinds of measuring instruments and also in such diverse applications as microwave tubes, toys and electronic watches. Although the quality of the materials for permanent magnets has been improved very considerably in the last 20 or 30 years, there is still a need to investigate the possibilities for further improvement. A survey is given in this article of the present state of knowledge of the physics of magnetic hardness, the main foundation on which this work is based.

Applications and general properties

The appearance of permanent-magnet materials such as 'Ticonal' [1] and ferroxdure [2] was followed by a great increase in the applications of the permanent magnet. Compared with electromagnets (including power supplies), permanent magnets offer the advantage of a larger ratio of the useful magnetic field energy to the volume of the magnet system. Their usefulness is of course particularly apparent where a constant magnetic field is required. As is widely known, the constancy of the externally generated field is related to the magnetic 'hardness' of the material, that is to say the extent to which the material retains its magnetization in opposing fields. In this way the polarization — and therefore the external field — of a permanent magnet is maintained. A particular example of an opposing field is the internal field of the poles of the magnet itself. In this case the demagnetizing action is again unable to destroy the polarization of the magnet. The present increasing interest in the further development of hard magnetic materials is explained in part by the growing demand for miniaturization in modern technology. The problem of heat dissipation is inseparable from miniaturization, and the substitution of permanent magnets for electromagnets obviously goes a long way towards solving that problem.

Dr H. Zijlstra is with Philips Research Laboratories, Eindhoven.

To ensure the most effective development it is desirable to start by investigating the likely applications of permanent magnets. The next step is to decide on the criteria that indicate suitability for these applications. These criteria can then provide a pattern for the production of tailor-made magnetic materials. This calls for insight into the effects that variation of such properties as remanence and coercivity has on the suitability of the materials for a particular application, and it also requires knowledge of the physical and technological (chemical) background. This will be the main subject of the present article.

The hardness of a magnetic material is due to the magnetic anisotropy exhibited by the crystallites of the material: each crystallite has one or more directions or axes of easy magnetization. It takes extra work to magnetize the crystallite in a direction other than the preferred or easy direction. The simplest case is a magnetic crystal with one axis of easy magnetization, and it is assumed that an amount of work E per unit volume is necessary to rotate the polarization vector J

[1] B. Jonas and H. J. Meerkamp van Embden, New kinds of steel of high magnetic power, Philips tech. Rev. 6, 8-11, 1941. 'Ticonal' is a trade name registered by N.V. Philips' Gloeilampenfabrieken, Eindhoven.

[2] J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Ferroxdure, a class of new permanent magnet materials, Philips tech. Rev. 13, 194-208, 1951/52.

through an angle Θ with respect to the axis. This anisotropy is described by an effective field H_a (called the anisotropy field), operating along the preferred axis, which attempts to keep the polarization vector oriented along this axis. For small angles of rotation the work is therefore given to a good approximation by [*]:

$$E = \tfrac{1}{2} H_a J \Theta^2.$$

The anisotropy field is a useful quantity for establishing stability criteria. If a crystal magnetized along its easy axis is placed in an increasing field in the opposite direction, instability will occur as soon as this field becomes equal to the anisotropy field. When the opposing field is increased further the polarization reverses and the crystal is then homogeneously magnetized in the direction of the applied field. In this situation the intrinsic coercive force, i.e. the field H_c needed for reducing the polarization J of the crystal to zero, is equal to the anisotropy field. It will be shown later that, for various reasons, instability does in fact occur with weaker opposing fields. In general H_c is weaker than the anisotropy field and only in very rare cases equal to it. In some materials the value of H_a is very high. A record value has been measured in the compound $SmCo_5$, in which H_a was equal to 2.3×10^7 A m^{-1} (i.e. about 3×10^5 oersteds) [3]. This material is therefore eminently suitable for making permanent magnets.

Table I lists various machines, devices and components in which permanent magnets are nowadays used [4]. The classification is based on four principles:
— mechanical energy is converted into electrical energy (or vice versa) in the magnetic field;
— the permanent magnet exerts a force on a ferromagnetically soft body;
— the permanent magnet is subjected to a directional force exerted by a magnetic field;
— the permanent magnet exerts a force on moving charge carriers, e.g. a beam of electrons in a vacuum.

In the next section the two main suitability criteria will be discussed that together cover almost the entire field of applications. They are the maximum energy product and the maximum change in the magnetic free energy [5]. Applications not covered by these criteria can be found among the positioning mechanisms in Table I. The subsequent four sections will deal with the various types of magnetic anisotropy and with the way in which they affect the magnetic hardness of the materials.

Apart from the fact that existing applications provide an incentive to search for better magnetic materials, the converse is of course equally true: better magnets lead to applications that had not previously been thought of or did not seem feasible. An example is to be seen in the recent experiments with magnetic levitation to avoid frictional forces in certain kinds of vehicle. The aim with these experiments is to make extremely fast tracked transport possible [6].

Table I. Examples of machines, devices and components using permanent magnets, classified by four functions that the magnet can perform.

Function	Application
Conversion of electrical into mechanical energy and vice versa	Small electric motors, dynamos, loudspeakers, microphones, eddy-current brakes, speedometers, magnetos
Exerting a force on a ferromagnetically soft body	Relays, couplings, bearings, clutches, magnetic chucks and clamps, separators (extraction of iron impurities, concentration of ores)
Alignment with respect to a field	Positioning mechanisms (e.g. stepping motors), compasses, some ammeters
Exerting a force on moving charge carriers	Magnetrons, travelling-wave tubes, some cathode-ray tubes, Hall plates

Two suitability criteria

The energy product

The extent to which a material will be suitable for applications in which electrical energy plays a part (the first group in Table I) depends on the amount of magnetic flux linkage per square metre and the maximum opposing field that can be tolerated without loss of polarization.

The *product* of the flux density B and the associated opposing field H, referred to as the energy product, is a useful measure of the performance of a particular magnet, since it is proportional to the potential energy of the field in the air gap. It is only useful, however, when the magnet is not disturbed by fields from another source. To determine the energy product it is of course necessary to have information about the hysteresis loop of the material (*fig. 1*). A permanent magnet

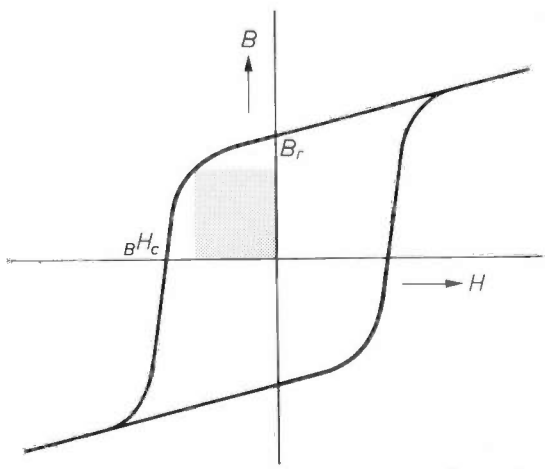


Fig. 1. Magnetic hysteresis loop; the shaded area is equal to the maximum energy product $(BH)_{max}$. B density of the magnetic flux. B_r remanence. H magnetic field-strength in the material. $_BH_c$ the B-coercive force.

that is subject only to the influence of its own field will be in a state represented by a working point in the second (or the fourth) quadrant of the hysteresis loop. In these quadrants the field is opposed to the flux density, and is referred to as the demagnetizing field.

It can be shown quite generally that the occurrence of a magnetic field *outside* the permanent magnet does in fact relate to a field inside the magnet with $\boldsymbol{B}$ and $\boldsymbol{H}$ in opposition. To do this, we have to apply Maxwell's equations to a situation in which there are no electric currents (apart from the circular currents on an atomic scale, which are the carriers of the magnetization of the material). The magnetic field-strength $\boldsymbol{H}$ then satisfies

$$\text{curl } \boldsymbol{H} = 0,$$

and for the flux density $\boldsymbol{B}$ we always have

$$\text{div } \boldsymbol{B} = 0.$$

For a permanent magnet of finite dimensions we may therefore deduce

$$\int_R (\boldsymbol{H} \,.\, \boldsymbol{B})\, \mathrm{d}V = 0,$$

where the integration is performed over the complete space R [7]. If this integral is written as the sum of the integral over the volume (R_{magn}) of the permanent magnet and the integral over the rest of the complete space (R_{rest}), then

$$\int_{R_{\text{magn}}} (\boldsymbol{H} \,.\, \boldsymbol{B})\, \mathrm{d}V = -\int_{R_{\text{rest}}} (\boldsymbol{H} \,.\, \boldsymbol{B})\, \mathrm{d}V.$$

Assuming that the space R_{rest} is 'empty', i.e. contains no magnetic substances, then the flux density there is given by $\boldsymbol{B} = \mu_0 \boldsymbol{H}$. The right-hand side of the last equation is then negative, which is only possible if $\boldsymbol{B}$ and $\boldsymbol{H}$ inside the magnet are of opposite sense or at least include an obtuse angle. This result is not affected if R_{rest} contains soft magnetic material in which $\boldsymbol{B}$ and $\boldsymbol{H}$ always have the same direction.

It can also be shown directly from what we have said above why the product BH is a good criterion of quality for the applications considered in this section. If we assume that any field present in soft magnetic material is negligible, we may write:

$$\int_{R_{\text{magn}}} (\boldsymbol{H} \,.\, \boldsymbol{B})\, \mathrm{d}V = -\mu_0 \int_{R_{\text{rest}}} H^2\, \mathrm{d}V.$$

The right-hand side of this equation is twice the potential energy of the field outside the magnet (i.e. in the air gap). This is proportional to $\boldsymbol{H} \,.\, \boldsymbol{B}$.

The exact location of the operating point — and hence the value of the energy product — depends on the relative dimensions of the magnet and the magnetic circuit in which it is used.

In the limiting cases of an infinitely long needle ($H = 0$) or of an infinitely extensive plate ($B = 0$) the energy product is equal to zero; there is then no external field. Between these two extremes a situation exists in which the energy product has its maximum magnitude. In the case of the needle-shaped magnet the demagnetizing field is very weak and the working point is close to the point B_r in fig. 1. The value of the flux density at this point is the remanence. If the magnet is made shorter and thicker, the working point then moves along the loop in the direction of the point ${}_BH_c$ (the value of H at this point is referred to as the B-coercive force), which it reaches if the magnet is given the form of a thin plate magnetized perpendicular to its plane. The demagnetizing field then has its maximum value and exactly compensates the magnetization. In a properly dimensioned design the energy product will thus assume a maximum value, $(BH)_{\max}$, *which is determined solely by the material used.* The suitability criterion sought has thus been found.

The product can be represented by the area of the shaded rectangle in fig. 1; its magnitude is equal to twice the total potential energy of the field produced outside the magnet, divided by the volume of the magnet. The higher the remanence, the greater the coercive force or the more convex the hysteresis loop, the greater is the value of the product. For an ideal magnet, i.e. a magnet that maintains the saturation value J_s of its polarization in spite of the presence of an opposing field H, the hysteresis loop in the second quadrant is formed by a straight line going from the point where $H = 0$, i.e. where $B = B_r = J_s$, to the point where $H = {}_BH_c = -J_s/\mu_0$. The maximum energy product is then given by:

$$(BH)_{\max} = \frac{1}{4\mu_0} J_s^2 .$$

To reach this maximum it is sufficient if the magnet maintains its saturation until the opposing field reaches the value $-\frac{1}{2}J_s/\mu_0$. A further improvement in the energy

---

[*] In this article SI units are used in all equations: the unit of magnetic flux density (magnetic induction) B is the tesla ($\mathrm{T} = \mathrm{Wb\,m^{-2}} = \mathrm{Vs\,m^{-2}}$), the magnetic field-strength H is in ampères per metre; the magnetic polarization is J ($= B - \mu_0 H$), the magnetization is M ($= B/\mu_0 - H$), μ_0 is the absolute permeability of free space ($= 4\pi \times 10^{-7}$ henries per metre).
In the Gaussian system of units the unit for B is the gauss and for H the oersted. The permeability of free space is then equal to 1, and the polarization J is given by $4\pi J = B - H$.

[3] K. H. J. Buschow and W. A. J. J. Velge, Z. angew. Physik **26**, 157, 1969. See also K. H. J. Buschow, W. Luiten, P. A. Naastepad and F. F. Westendorp, Philips tech. Rev. **29**, 336, 1968.

[4] See also F. G. Tyack, Permanent magnet applications, in: D. Hadfield (ed.), Permanent magnets and magnetism, Iliffe Books Ltd., London 1962, pp. 297-372.

[5] The magnetic free energy is a quantity from the thermodynamics of magnetic systems. When a reversible change of state takes place in a system at constant temperature, the amount of mechanical work that the system then freely exchanges with the environment is called the free energy of the system.

[6] The levitating of permanent magnets involves stability problems, for which a solution is given by H. van der Heide in Stabilization by oscillation, Philips tech. Rev. **34**, 61-72, 1974 (No. 2/3).

[7] The derivation is given in: W. F. Brown, Jr., Magnetostatic principles in ferromagnetism, North-Holland Publ. Co., Amsterdam 1962, p. 44.

product is then only possible with materials that have a higher saturation value J_s. The highest known saturation value at room temperature is shown by an FeCo alloy (2.4 Wb m^{-2}); from this value the theoretical energy product could be as much as 1150 kJ m^{-3}. However, the coercive force of the alloy is very low, which makes the material unsuitable for permanent magnets.

Fig. 2 shows the improvements achieved in maximum energy products over the years, the record values being indicated on a logarithmic scale. It is interesting to note how closely the curve approximates to an exponential development.

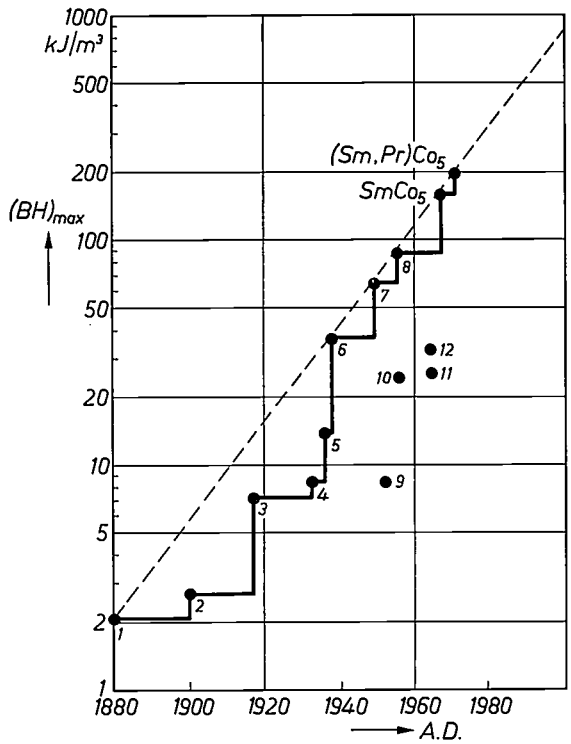


Fig. 2. Historical trend of the maximum energy products $(BH)_{max}$ achieved experimentally since the year 1880. The dashed line represents an exponential increase that corresponds reasonably well, particularly since 1940, with the reality. *1* C steel. *2* W steel. *3* Co steel. *4* Fe-Ni-Al alloy. *5* 'Ticonal II'. *6* 'Ticonal G'. *7* 'Ticonal GG'. *8* 'Ticonal XX' (laboratory [12]). *9* ferroxdure 100. *10* ferroxdure 300. *11* ferroxdure 330. *12* ferroxdure 360.

Magnetic free energy

In applications involving clamping ability, lifting power or pull of the magnet (ponderomotive force, the second category in Table I) the working point is also in the second quadrant of the hysteresis loop. Whereas in the previous group of applications it was the *location* of the working point that mattered, the important thing now is how the working point moves. If, for example, the application is of a cyclical nature, it is usually necessary for the working point to 'stay well on the loop' during the cyclical motion, so that good reversibility is important. The amount of mechanical work done in going anticlockwise round part of the loop and completely regained on going back again is used as a criterion for measuring the performance of a magnet system for applications of this type.

In these applications there is generally a particular configuration of permanent magnets and magnetizable objects, which are capable of relative movement. Leaving aside the work required to overcome friction, the mechanical work required to produce an isothermal change in the configuration is equal to the increase in its magnetic free energy. Conversely, a decrease in the magnetic free energy will result in the same amount of mechanical work becoming available.

According to the first law of thermodynamics (conservation of energy) a system in which a reversible process takes place can be described by the equation

$$T\mathrm{d}S + \mathrm{d}A = \mathrm{d}U.$$

The term $T\mathrm{d}S$, the product of the absolute temperature T of the system and the change of its entropy S, is equal to the amount of heat supplied to the system from the environment. In addition the environment performs on the system an amount of mechanical work $\mathrm{d}A$, taken as positive. This sign convention for the mechanical work performed is employed for systems in which magnetic effects occur. Both amounts of energy are spent on the increment $\mathrm{d}U$ of the internal energy of the system.

The free energy F of the system is defined by:

$$F = U - TS.$$

It follows from these two relations that

$$\mathrm{d}F = \mathrm{d}A - S\mathrm{d}T.$$

If the state of the system changes isothermally (i.e. $\mathrm{d}T = 0$). then

$$\mathrm{d}F = \mathrm{d}A.$$

In a system that contains magnetic material the main problem is to find the correct expression for the mechanical work.

The criterion used for the suitability of a magnetic material for applications of the type we are now considering is the maximum possible reversible change of its magnetic free energy. This value is usually calculated per unit volume of the magnet.

The mechanical work $\mathrm{d}A$ associated with an infinitesimal change of the configuration is equal to

$$\mathrm{d}A = \tfrac{1}{2}\int_{R_{magn}} (\boldsymbol{H} \,.\, \mathrm{d}J - J \,.\, \mathrm{d}\boldsymbol{H})\mathrm{d}V,$$

where the integration is performed over the part R_{magn} of the space occupied by the material. To show that this expression for the mechanical work is reasonable [8], let us imagine a number of bodies of various magnetizations and arranged in a particular configuration. We assume that the bodies are situated in each other's

magnetic fields and that their temperature remains constant. A slight change in the configuration causes a change in the fields and hence in the polarizations. For each body the increase in the magnetic free energy consists of a quantity dF_p — connected with the build-up of the polarization in the material — and a quantity of interaction energy dF_i — since in a field $\boldsymbol{H}$ a piece of material with the polarization vector $\boldsymbol{J}$ possesses the potential energy $-(\boldsymbol{J} . \boldsymbol{H})$. For the body considered we can now write:

$$dF = dF_p + dF_i = \boldsymbol{H} . d\boldsymbol{J} - d(\boldsymbol{J} . \boldsymbol{H}).$$

To find the change in the free energy of the whole system we must perform a summation over all the bodies. The contributions from the interaction energy would then be counted twice, but putting a factor $\frac{1}{2}$ in front of them corrects for this.

The total increase in the free energy is therefore

$$dF_{system} = \sum dF_p + \tfrac{1}{2} \sum dF_i ,$$

where both summations are made over all the bodies. Using the above expressions for dF_p and dF_i and applying the expression $dF = dA$ for isothermal changes, we obtain the required expression for the mechanical work, calculated per unit volume of the material.

We should note here that the energy change $\boldsymbol{H} . d\boldsymbol{J}$ is positive, because the structure of the material offers a certain resistance to the change of the polarization. The interaction energy $-(\boldsymbol{J} . \boldsymbol{H})$ has a minus sign because it is customary to take this energy by definition equal to zero for two bodies that are an infinite distance apart.

If the polarization vector in the expression for the mechanical work is replaced by the equivalent quantity $\boldsymbol{B} - \mu_0 \boldsymbol{H}$, then, after integration,

$$dA = \tfrac{1}{2} \int_{R_{magn}} (\boldsymbol{H} . d\boldsymbol{B} - \boldsymbol{B} . d\boldsymbol{H}) dV .$$

To evaluate this integral it is necessary to bear in mind that during a change in the configuration the working point moves along the hysteresis loop in the second quadrant from P to Q (*fig. 3*). It is then found that the work dA is equal to the area of the sector OPQ. One configuration (point P) cannot move farther to the right than point B_r, where H is zero and therefore the magnetic circuit must be closed. The other configuration (point Q) cannot move farther to the left than the point $-{}_BH_c$, where B and the force exerted are zero. Where possible, cyclical processes will be carried out in such a way that the working region in the second quadrant extends to the vertical axis (B_r). The magnetic circuit there is closed, which corresponds to a state of lowest energy. In general the material chosen for these applications is one in which the working point can move *reversibly* from remanence over the greatest possible extent of the hysteresis loop. For an ideal magnet, where the complete (linear) hysteresis branch in the second quadrant is traversed reversibly, the maximum mechanical work made available per unit volume during a change of configuration is given by:

$$\tfrac{1}{2} B_r \, {}_BH_c = \frac{1}{2 \mu_0} J_s^2 .$$

It will be evident that the magnet must be capable of maintaining its saturation polarization J_s until the opposing field reaches the value $-J_s/\mu_0$. This places a much more difficult requirement on the hardness than in applications for which the first criterion applies ($-\frac{1}{2} J_s/\mu_0$).

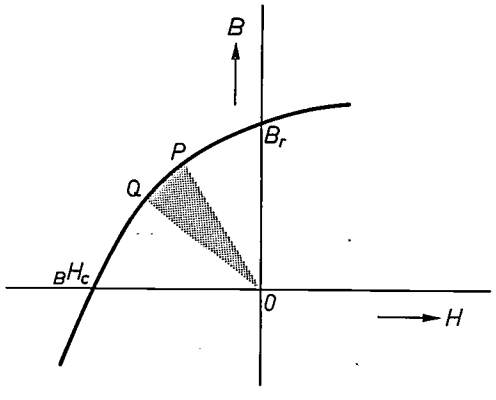


Fig. 3. Part of a hysteresis curve in the second quadrant; B_r the remanence, ${}_BH_c$ the B-coercive force. The area of the sector OPQ is equal to the mechanical work that has to be performed on the system (e.g. a permanent magnet near a soft-iron object) to change the configuration corresponding to point P to the configuration of point Q.

Fig. 4 shows the trend in the development of the maximum amount of mechanical work that can be obtained reversibly from a well designed configuration of a magnet and a magnetizable object. The amounts of work are given per cubic metre of permanent magnet. It can be seen that the rise resembles that in fig. 2, but there are some striking differences. For example, in fig. 4 the material $SmCo_5$ gives a better result than $(Sm,Pr)Co_5$, whereas in fig. 2 the opposite was the case. The explanation for this is that the coercive force of $SmCo_5$ is greater than that of the other material. The rapid increase after about the year 1945 — beginning with the introduction of 'Ticonal G' and ferroxdure — is also connected with the higher coercivity obtained. This has made the newer materials ideally suitable for 'dynamic' applications, where the important force is the attractive or repulsive force.

[8] See R. M. Bozorth, Ferromagnetism, Van Nostrand, New York 1951, pp. 729-731.

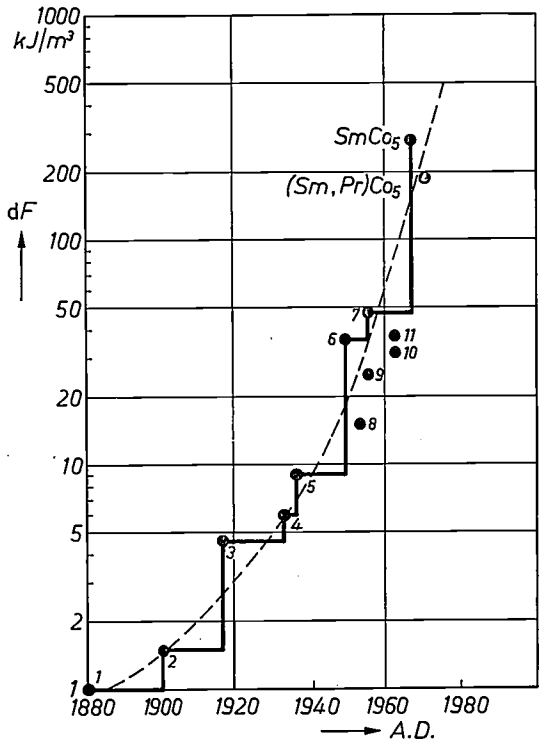


Fig. 4. Historical trend of the maximum reversible change of the magnetic free energy, dF, achieved since the year 1880. *1* C steel. *2* W steel. *3* Co steel. *4* Fe-Ni-Al alloy. *5* 'Ticonal II'. *6* 'Ticonal GG'. *7* 'Ticonal XX' (laboratory [12]). *8* ferroxdure 100. *9* ferroxdure 300. *10* ferroxdure 360. *11* ferroxdure 330. The dashed line gives an impression of the remarkably steep rise.

In addition to the criteria we have been looking at there are of course others for a complete assessment of magnetic materials. These include chemical and mechanical stability, electrical resistance and a variety of characteristics underlying their workability. Not much will be said about these, however, in this article. What we are primarily concerned with are the physics of the remanence, the coercivity and the shape of the hysteresis loop (especially in the second quadrant). We shall therefore now turn to a discussion of the various types of magnetic anisotropy, each of which is associated with its own kind of magnetic hardness.

Shape anisotropy

Shape anisotropy refers to the preference that the polarization in a long body has for the direction of the major axis. The effect, which does not arise from an intrinsic property of the material, can easily be described in the case of a prolate ellipsoid. It is assumed that the ellipsoid is homogeneously magnetized in a direction that makes an angle Θ with the major axis (*fig. 5*). The demagnetizing field H_d due to the magnetic poles at the surface is also homogeneous within the ellipsoid. Along each of the three principal axes of the ellipsoid we can apply one of the relations

$$\mu_0 H_{di} = -N_i J_i,$$

where i is the number of the principal axis; the coefficients N_i are the demagnetization factors. In the case of an ellipsoid of revolution we have the 'parallel' demagnetization factor $N_{||}$ for the direction parallel to the axis of revolution, and two 'perpendicular' demagnetization factors $N_\perp$, which, for reasons of symmetry, are identical.

Using the equation quoted earlier,

$$\int_R (\boldsymbol{H} \,.\, \boldsymbol{B}) \mathrm{d}V = 0,$$

it can easily be shown that the energy E_m of the demagnetizing field, which is given by

$$E_m = \tfrac{1}{2}\mu_0 \int_R H^2 \mathrm{d}V,$$

depends in the following way on the parameters that describe the situation:

$$E_m = \frac{1}{2\,\mu_0}\left\{N_{||}J^2 + (N_\perp - N_{||})J^2 \sin^2\Theta\right\} R_{magn}.$$

In this expression R_{magn} is the volume of the ellipsoid. The coefficient of the directionally dependent part of the equation describes the shape anisotropy. For small values of the angle Θ we find from this an anisotropy field

$$H_a = (N_\perp - N_{||})J/\mu_0.$$

In the case of a long bar (needle) this field H_a approximates to the value $J/2\mu_0$. If the bar consists of saturated

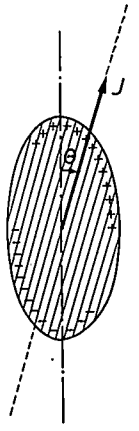
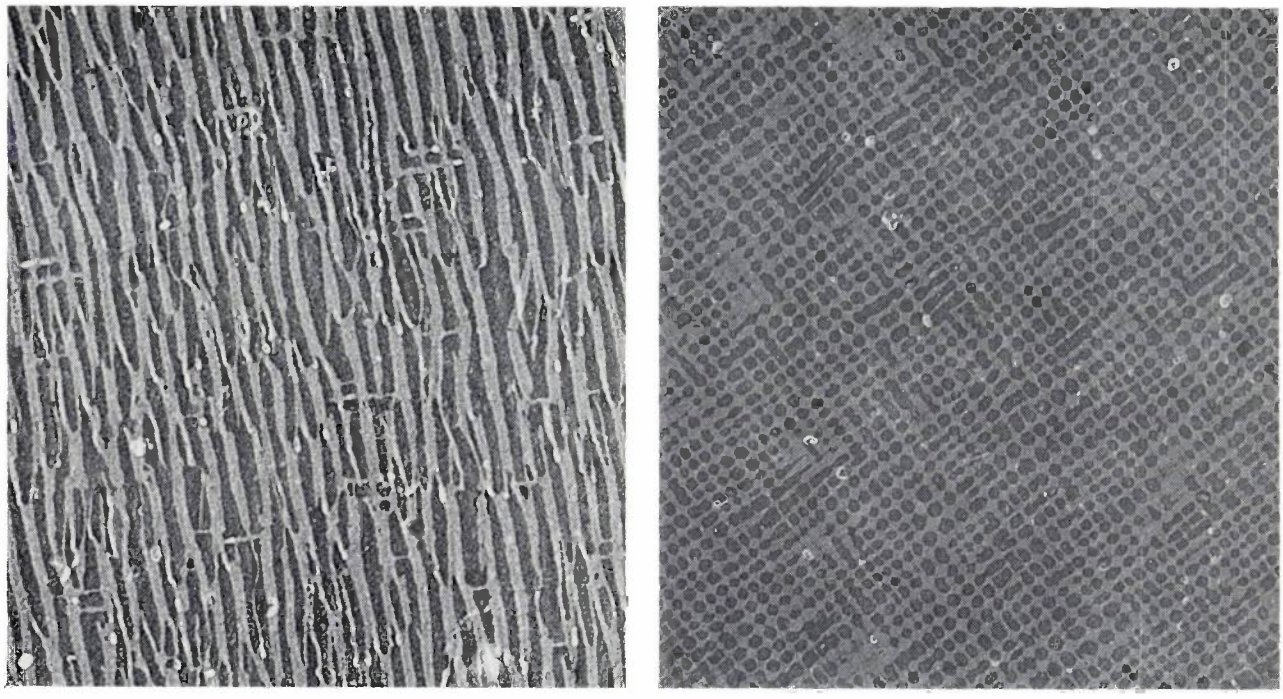


Fig. 5. A magnetized ellipsoid of revolution of a ferromagnetic material, where the polarization vector J makes a small angle Θ with the major axis. The demagnetizing field inside the ellipsoid is homogeneous.



a *b*

Fig. 6. Example of a fine dispersion of magnetic needles (mainly FeCo) in a matrix (NiAl) that has a much smaller magnetic moment. *a)* Plane of observation parallel to the easy direction of magnetization. *b)* The same, but perpendicular to the easy direction. (Electron micrograph of 'Ticonal XX' magnet steel; from K. J. de Vos, thesis, Delft 1966.) Magnification 50 000×.

iron ($J_s \approx 2$ Wb m^{-2}) it follows from the foregoing that $H_a \approx 10^6$ A m^{-1} ($\approx 10^4$ oersteds). This is the value that the coercive force has when the magnetization is rotated uniformly (coherently).

Magnetic material that derives its hardness from its shape anisotropy consists of a fine dispersion of magnetic needles in a matrix of non-magnetic or weakly magnetic material (*fig. 6*). In magnets of such material the coercive force is in general much smaller than that which we have just derived for a single needle. One of the reasons for this is the partial compensation of the demagnetizing field, which is due to the magnetic interaction between the needles. However, the coercive force may also be smaller in individual needles, an effect which is due to incoherent rotation of the magnetization ('curling'). Since it also occurs in needles contained in a matrix, this effect has a similarly adverse influence on the coercivity of the magnetic material. We shall now take a somewhat closer look at this effect of incoherent rotation and at the compensation of the demagnetizing field.

Incoherent rotation

Depending on the thickness of a ferromagnetic bar (needle) a reversal process is possible in which the magnetization does not remain homogeneous during the process, implying a reduction of the magnetostatic energy barrier. Although a certain ferromagnetic spin-coupling energy must be produced where such non-uniform (incoherent) rotation of the magnetization occurs, that energy decreases as the cross-section of the bar increases. The process was predicted on theoretical grounds in 1957 [9]; its experimental confirmation, in Fe and FeCo needles, dates from 1964 [10]. An intermediate state of the process is illustrated in *fig. 7*. It can be seen that in the ferromagnetic bar each atomic moment takes up a particular oblique position in the local plane tangential to the cylindrical surface containing the carrier of that moment. This angular position is zero exactly on the axis of the bar, in other words there is no change in the magnetic vector at that position. Farther away from the axis the change in the angular position becomes greater. For each cylindrical surface the angular position has a particular value. There is also a cylindrical surface inside the bar where the angular change amounts to $\pi/2$ radians, so that the atomic vectors there combine to give a purely tangential state of magnetization. At a greater distance from the axis of the bar the change in the angular position is greater than $\pi/2$, which increasingly comes to resemble

[9] E. H. Frei, S. Shtrikman and D. Treves, Phys. Rev. **106**, 446, 1957.

[10] F. E. Luborsky and C. R. Morelock, J. appl. Phys. **35**, 2055, 1964.

a 'real' reversal. In the situation illustrated no lines of force emerge from the outside surface of the (infinitely long) cylinder. The energy barrier that opposes the reversal of the magnetization does not therefore contain magnetostatic energy; in fact the barrier consists only of a ferromagnetic spin-coupling energy.

In general it would appear that it is best to make the packing density as high as possible, to improve the flux density of the magnet. However, it is at the very high packing densities that the compensation of the demagnetizing field, which is the second cause of too low a coercive force, becomes particularly apparent.

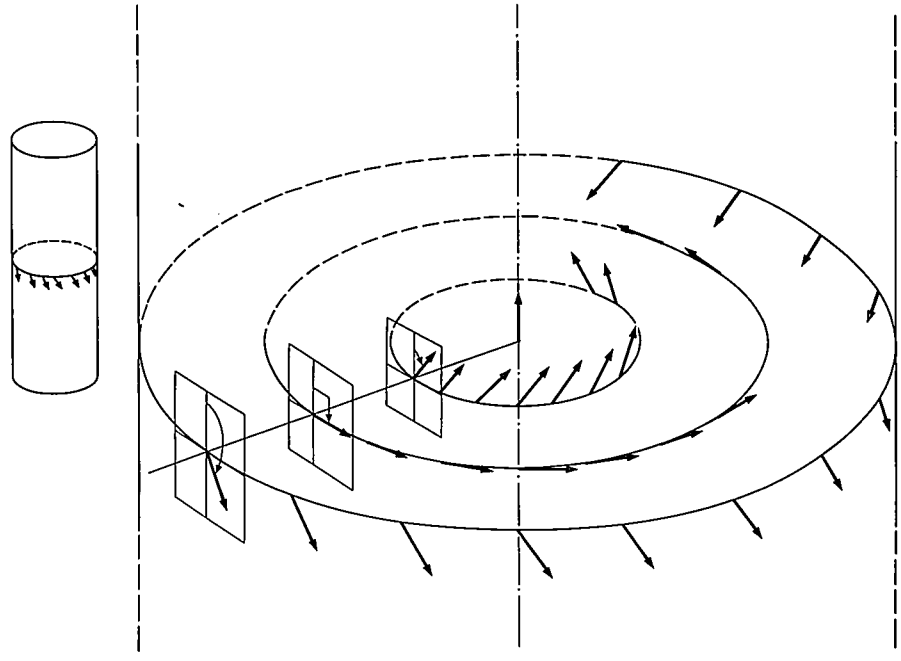


Fig. 7. Intermediate state occurring in a non-uniform magnetization reversal mechanism (called 'curling') in an infinitely long cylinder of circular cross-section.

In *fig. 8* the coercive force (here the critical field-strength at which the originally homogeneous axial magnetization becomes unstable) is plotted as a function of the bar diameter, both for the case of 'curling' and for ordinary, uniform rotation. Above a critical diameter, which is about 17 nm for the Fe and FeCo needles, the non-uniform rotation is energetically more favourable, and will thus determine the coercive force. This explains why the coercive force found is lower than the maximum $J_s/2\mu_0$, the value for uniform rotation.

The situation for the complete magnet is not very different, since for curling at least the external magnetostatic field of the needles is zero and there is therefore no interaction. As long as this is the case, it is found that the coercive force of the complete magnet is in fact lower than the maximum $J_s/2\mu_0$, and what is more the coercive force would then be expected to be independent of the packing density of the needles.

In fact the situation is more complicated: the expected independence of the packing density does not appear to exist. The reason for this is that the increasing packing density makes the ordinary, uniform rotation mechanism more readily possible, until finally non-uniform rotation vanishes completely.

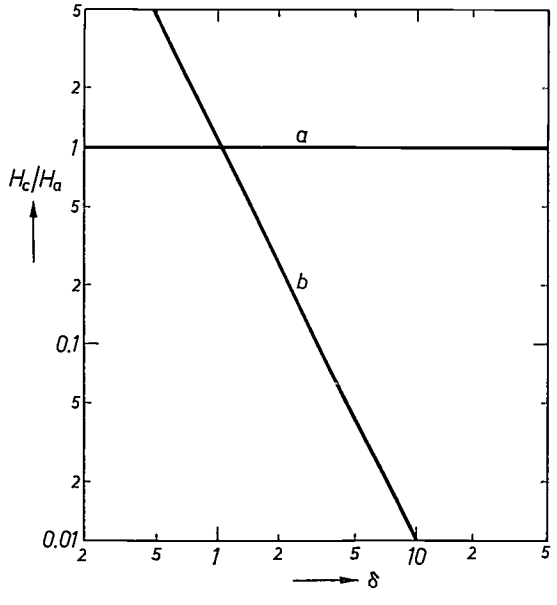


Fig. 8. The calculated coercive force (curve *a*) in the normal, uniform demagnetizing situation and (curve *b*) in the 'curling' situation of fig. 7, as a function of the diameter of the magnetized cylinder. The normalized coercive force H_c/H_a is plotted as a function of the normalized diameter δ. This is equal to $2rJ_s/(2\mu_0 A^{\frac{1}{2}})$, where A is the coupling constant between two neighbouring spins. $2r$ real diameter. H_a strength of the anisotropy field. J_s saturation value of the polarization. μ_0 permeability of free space.

Compensation of the demagnetizing field

If a magnet whose hardness is derived from the shape anisotropy of closely packed needles is demagnetized by uniform rotation, strong magnetic interaction will occur between the needles during that process. As a result the energy barrier to uniform rotation of the magnetization of a needle is considerably lowered, because the field of the surrounding needles partly compensates the field of that needle. The coercive force in this magnet will thus be smaller than the theoretical maximum for a single needle.

A limiting case of this occurs when the packing density is increased so far that the situation is approached in which the needles are no longer separately distinguishable. In this case the effective shape anisotropy — and hence the coercive force — will decrease until both quantities finally become zero when this total packing density is reached. The lowering of the energy barrier as just described corresponds to a downward shift of curve *a* in fig. 8, which makes the point of intersection move to larger diameters.

A magnet based on shape anisotropy that is widely used in technical applications is the 'Ticonal' magnet [1]. The material 'Ticonal' is an alloy of Fe, Co, Ni, Al and Cu, sometimes with the addition of some Ti. This alloy is subjected to a heat treatment in a magnetic field so as to bring it into a state in which ferromagnetic FeCo needles of about 30 nm thick are finely dispersed in a non-magnetic matrix (see fig. 6). *Fig. 9* shows the second quadrant of the hysteresis loop. The loop is practically rectangular. The coercive force is about one-third of the effective anisotropy field, as measured by a torsion magnetometer [11]. The conclusion to be drawn from this is that a mechanism of non-uniform rotation determines the coercive force in this case. The mechanism may be connected with the occurrence of curling, but it could also be due to irregularities in the distribution and shape of the needles, since such irregularities can initiate the reversal process locally. The hysteresis loop mentioned here relates to a laboratory-made single crystal of 'Ticonal' with a maximum energy product of 90 kJ m^{-3} ($\approx 11 \times 10^6$ GOe) [12]. In the industrial product the value of $(BH)_{max}$ is usually somewhat lower.

Crystal anisotropy

Three forms

There are three situations that give rise to magnetic anisotropy as an intrinsic crystal property. The first and most important one is that in which the atoms possess an electron-orbital moment in addition to an electron-spin moment. In such a situation the spin direction may be coupled to the crystal axes. This arises through the coupling between spin and orbital moments and the interaction between the charge distribution over the orbit and the electrostatic field of the surrounding atoms. There will then be one or more axes or surfaces along which the magnetization requires relatively little work. The crystal will then be preferentially magnetized along such an easy axis or plane.

The second situation is encountered in non-cubic crystal lattices. In these crystals the magnetostatic interaction between the atomic moments is also anisotropic, which may give rise to easy directions or planes of magnetization.

The third possibility of crystal anisotropy is found in the directional ordering of atoms as described by L. Néel [13]. This typically involves solid solutions of atoms of two kinds A and B, linked by the atomic bonds A-A, A-B and B-B. In the presence of a strong external magnetic field the internal energy of these bonds may be to some extent direction-dependent. Given a sufficient degree of atomic diffusion — as a result of raising the temperature, for example — a certain ordering can be brought about in the distribution of the bonds; in this way it is possible to 'bake' the direction of this field into the material as the easy axis of magnetization.

An example of a permanent-magnet material based on uniaxial crystal anisotropy is ferroxdure [14]. This material consists of the compound $MO(Fe_2O_3)_6$, where

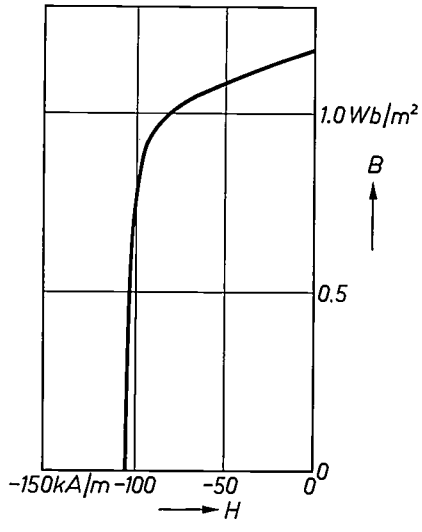


Fig. 9. The second quadrant of the hysteresis loop of a single crystal of 'Ticonal' magnet steel [12]. The coercive force in this case is probably determined by the 'curling' mechanism (fig. 7).

[11] See H. Zijlstra, Experimental methods in magnetism, North-Holland Publ. Co., Amsterdam 1967.
[12] A. I. Luteijn and K. J. de Vos, Philips Res. Repts. **11**, 489, 1956.
[13] L. Néel, J. Phys. Radium **15**, 225, 1954.
[14] See the article of note [2]. An application in loudspeaker magnets, using both ferroxdure and 'Ticonal' to best advantage, was described by M. F. Reynst and W. T. Langendam in Philips tech. Rev. **24**, 150, 1962/63.

M is a divalent metal ion (Ba, Sr or Pb). The coercivity of this material is high (about five times higher than that of 'Ticonal G'), but its remanence is relatively low. The material has the advantage of possessing a particularly high electrical resistance, which makes it useful for many high-frequency applications. An additional advantage over 'Ticonal' is the absence of comparatively rare elements such as Co and Ni.

The magnetization processes

In the rest of the discussion we shall confine ourselves to crystals with only one easy axis. The anisotropy field of such crystals may often be very strong compared with the values encountered in the case of shape anisotropy. The crystals in the compound $SmCo_5$ mentioned above are so far the best example of this.

If the homogeneously magnetized crystal is situated in an increasing and opposing field, we should expect that all the atomic moments would reverse their direction simultaneously as soon as this field reaches the value H_a. This uniform rotation process has in fact been observed, for example in perfect iron crystals [15]. In the great majority of cases, however, the reversal of magnetization is found with fields much smaller than H_a, as in ferroxdure, where the coercive force is about 0.3 H_a. The highest value of coercive force thus far observed was found in $SmCo_5$, amounting to about 5×10^6 A m^{-1} (6×10^4 Oe) [16]; this value is only one-fifth of the anisotropy field. In nearly all materials that owe their magnetic hardness to crystal anisotropy, H_c/H_a ratios of the order of 0.1 are found.

The reversal of magnetization in these materials is not due to uniform rotation but to the dividing up of the originally homogeneously magnetized crystal into domains, each of which is oppositely magnetized with respect to the other. Between these domains (often called Weiss domains) there is a transitional zone, called the domain or Bloch wall, in which the orientation of the magnetization gradually changes from the direction of one domain to that of the neighbouring domain. The wall comprises both ferromagnetic coupling energy (the atomic moments in the wall are not oriented parallel to one another) and anisotropy energy (the atomic moments in the wall are not directed along an easy axis).

To obtain such a division into domains there must of course be at least one wall produced in the homogeneously magnetized crystal. The energy for creating the wall is supplied by the opposing field, which must reach a threshold value called the nucleation field-strength H_n before a wall can be formed. At this value H_n at least one atomic moment will then become unstable. The question now is: what determines the nucleation field-strength?

To answer this question we consider first a perfect crystal which is magnetically saturated and is located in an external field parallel to the easy axis. An individual atomic moment then has the following four effective fields acting upon it: the anisotropy field H_a, the Weiss field H_w (representing the coupling between the relevant moment and its nearest neighbours), the demagnetizing field H_d (due to the finite dimensions of the crystal) and the externally applied field H.

Instability will arise as soon as the field-strength of the externally applied field predominates over $H_a + H_w + H_d$; it thus follows that the nucleation field-strength H_n is equal to $-(H_a + H_w + H_d)$. The Weiss field is of the order of 10^9 A m^{-1} (about 10^7 Oe). The demagnetizing field is negative, and its maximum magnitude is of the order of 10^6 A m^{-1} (about 10^4 Oe); it can be neglected in comparison with H_w. An increasing opposing field will thus have to exceed the value $H_a + H_w$ to cause the instability of one atomic moment. However, as soon as the opposing field becomes equal to the anisotropy field H_a, *all* the atomic moments will then be capable of reversing *as a whole*. These considerations thus lead to the conclusion that the non-uniform processes are excluded in the case of perfect crystals, and that for these crystals the coercive force H_c is equal to the anisotropy field H_a. The fact that in experimental crystals H_c is usually smaller than H_a is attributable to lattice defects, which can reduce the coercive force through both H_a and H_w.

Lattice defects impair the coercive force because they reduce the value of H_a locally by disturbing the symmetry of the environment.

It should be noted here that the value of H_a can differ for atoms at the crystal surface from the values for atoms situated within the bulk of the crystal. The coupling between the atomic moments averages out the effect of these differences to an effect on the crystal anisotropy that can only have any significance in extremely thin films or in very small particles (about 10 nm) [13]. Experiments such as those by R. W. DeBlois [15] suggest that it is reasonable to assume that this effect is not likely to have any marked influence on the nucleation field-strength. In this context the study made by L. Liebermann *et al.*[17] may be relevant. They found that the magnetic moment of atoms close to the surface in thin films of Fe and Ni is smaller than that of the atoms below the surface. Although the effect only extends over a few atomic layers, and will therefore be barely perceptible in the magnetization of larger specimens, it could well have some influence on the creation of walls. It is difficult to predict what this influence would be; a lowering of the coupling energy would favour the creation of a wall, whereas lowering of the magnetic moment would have exactly the opposite effect of reducing the influence of an external field. Measurements on very thin layers of materials with a large magnetic anisotropy could provide useful information here.

What is presumably more important than the adverse effect of lattice defects on the anisotropy field is the

effect on the strength of the Weiss field. This field is highly sensitive to the distance between neighbouring atomic moments. When this distance is reduced the coupling may even acquire a negative sign, which would tend to give rise to antiparallel alignment. In such a situation it will be much easier to change the orientation of an atomic moment so that a domain wall could easily be initiated in an opposing field still definitely weaker than the anisotropy field H_a. An example is the 'τ phase' in the Mn-Al system [18]. In this compound stacking faults in the ordered crystal, in which such small distances are found, will very probably facilitate the creation of walls, even in a positive field [19]. A relatively low coercive force is then found ($H_c = 4\times10^5$ A m^{-1}, while $H_a = 3.2\times10^6$ A m^{-1}). If plastic deformation of an MnAl crystal introduces many stacking faults — on which walls will arise — the remanence will also be low relative to the saturation value of the polarization (*fig. 10*).

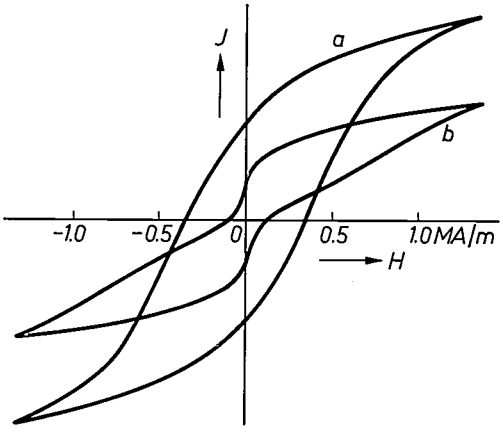


Fig. 10. Hysteresis loops of the τ phase in the system Mn-Al. Curve *a*: no deformation. Curve *b*: strong plastic deformation. In the second case there are antiparallel couplings between neighbouring atomic moments. Domain walls can then be created easily; the remanence is small compared with the saturation magnetization.

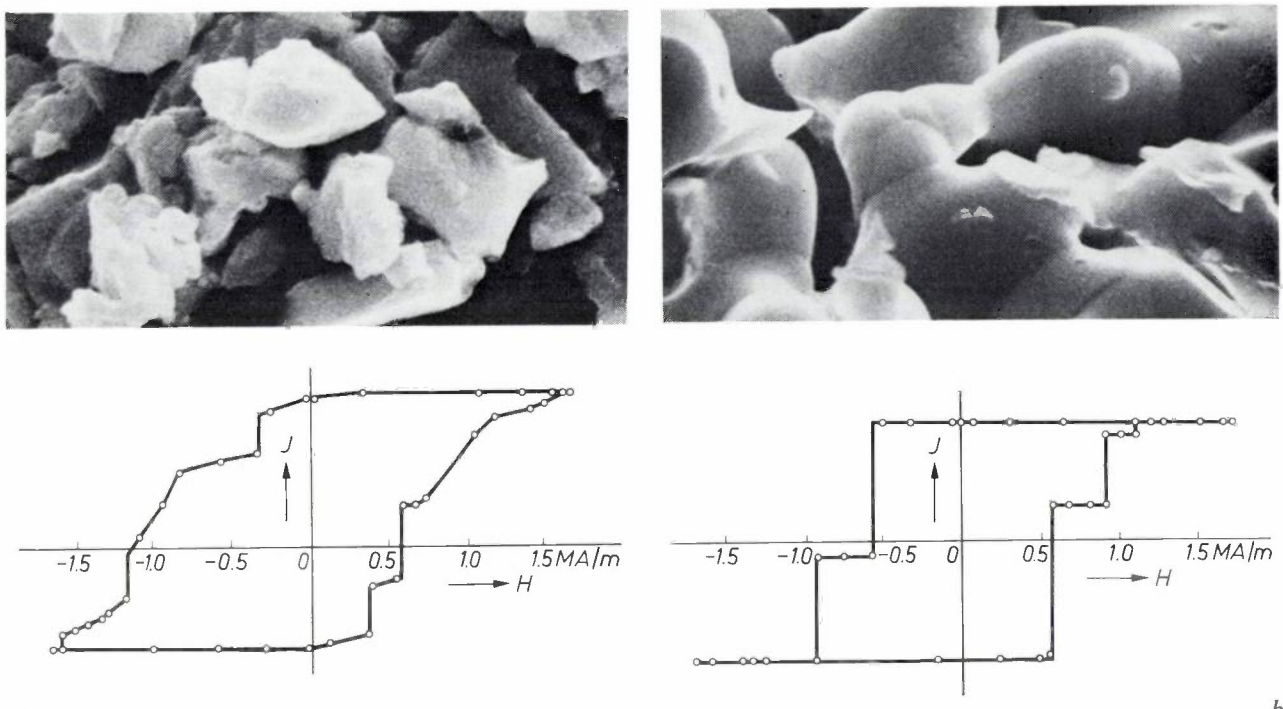
The coercive field is not zero because the same stacking faults tend to pin the domain walls. Walls tend to form at places where they possess lower energy than would be possible elsewhere; when a wall is present at such a place, correspondingly more work is required to displace it. The force which this requires is given by the 'pinning field' H_p. The pinning of domain walls by stacking faults in MnAl has been directly observed by Lorentz microscopy [20].

So far we have only been considering imperfections inside the lattice. However, the crystal surface itself may also have a considerable influence on the creation of walls. Near scratches and also at sharp edges strong field concentrations are present that will locally modify the direction of the magnetization. This field-strength, which in theory is infinitely high at infinitely sharp edges, can thus initiate the creation of walls. In practice, of course, the sharpness of the edge and hence the local field concentration is limited by the atomic structure.

The way in which the reversal of magnetization can be affected by the condition of the surface is illustrated in *fig. 11*. This relates to two $SmCo_5$ powders of comparable coercivity. The powders were made by milling the material. One of the powders was then heated to a temperature high enough to round off the sharp contours of the grains. The hysteresis loop of one grain of both powders was measured with a highly sensitive magnetometer specially designed for the purpose [21]. This method of observation makes it possible to distinguish individual Barkhausen jumps in the loop, which are identifiable with the creation and displacement of walls. Wall creation readily occurs in the powder with sharp-edged grains. The coercive force in this case is determined by the strength of the wall pinning. The powder with the rounded grains, on the other hand, opposes wall creation to such an extent that the pinning field no longer plays any part. As soon as a wall is formed, it passes rapidly through the grain and vanishes. In this case the coercive force is entirely determined by the nucleation field-strength. (The steps in the curve *b* are presumably connected with the presence of more than one grain; it is often difficult to avoid this.)

The extent to which wall creation and pinning can vary in character is neatly illustrated in *fig. 12*, which shows the hysteresis loop of a grain of $SmCo_5$ with a diameter of about 5 μm. The grain was first magnetized in a field of 4×10^6 A m^{-1} (about 5×10^4 Oe). At a field-strength of 0.6×10^6 A m^{-1} a wall is seen to form, which immediately passes through the greater part of the grain. In this case the coercive field is therefore equal to the nucleation field-strength. A slight increase in the opposing field is needed to remove the wall from the grain. The lower branch of the loop was measured after premagnetizing in a field of -4×10^6 A m^{-1}. The jump appears again at almost the same coercive force (= nucleation field-strength). However,

[15] R. W. DeBlois and C. P. Bean, J. appl. Phys. 30, 225 S, 1959.
[16] F. J. A. den Broeder, personal communication.
[17] L. Liebermann, J. Clinton, D. M. Edwards and J. Mathon, Phys. Rev. Letters 25, 232, 1970.
[18] A. J. J. Koch, P. Hokkeling, M. G. van der Steeg and K. J. de Vos, J. appl. Phys. 31, 75 S, 1960.
[19] H. Zijlstra, Z. angew. Physik 21, 6, 1966.
[20] H. Zijlstra and H. B. Haanstra, J. appl. Phys. 37, 2853, 1966. See also the article by the same authors in Philips tech. Rev. 29, 218, 1968.
[21] H. Zijlstra, A vibrating-reed magnetometer for microscopic particles, Philips tech. Rev. 31, 40-43, 1970.



a b

Fig. 11. Effect of the surface condition of $SmCo_5$ grains on the nature of the magnetic hysteresis. *a*) Angular grains produced by milling; the hysteresis is determined by pinning of domain walls. *b*) Grains that have been rounded by heat treatment at 1100 °C; the hysteresis is determined by the creation of walls.

when the field is reduced from about $+1.6\times10^6$ A m^{-1} it is found that a wall has remained behind in the grain. This wall is de-pinned at -0.33×10^6 A m^{-1}, and is then able to move through a large part of the grain with the application of very little force. A proof of this free mobility within the approximately spherical grain is the fact that the width of the relevant loop (the inner loop in the figure) is almost zero and that the slope is in good agreement with the theoretically predicted slope in a ferromagnetic sphere with freely mobile walls. This slope is then determined solely by the demagnetization factor. Obviously the pinning sites do not pin the wall strongly, and they lie mainly at the surface of the grain. It may well be in this case that wall creation and pinning take place at the same sites.

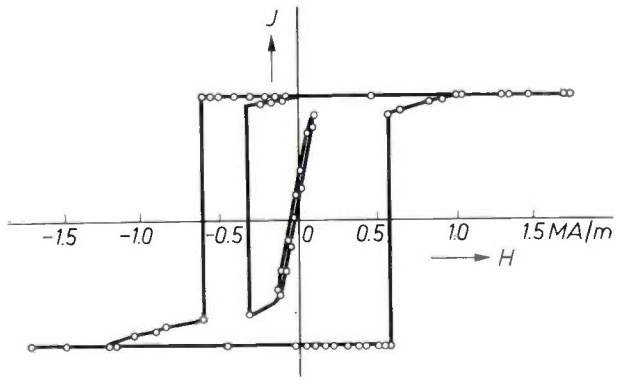


Fig. 12. Hysteresis loop of a grain of $SmCo_5$ (diameter about 5 μm). The coercive force is determined by wall creation, which requires a field of -0.6×10^6 A m^{-1} (about −7500 Oe). The narrowness of the inner loop indicates that a much weaker field is sufficient to move a wall from its pinning site, so that a wall once created can readily move through a large part of the grain.

The conclusion to be drawn here is that a magnet whose coercive force is determined by the creation of a wall is more suitable for a given application than a magnet whose coercive force is of the same magnitude but is determined by a pinning field. In the first case the magnetization is more strongly maintained against the opposing field; the hysteresis loop therefore encloses a larger area in the second quadrant, which results in a higher energy product $(BH)_{max}$.

Briefly summarizing what has been said in the foregoing about the factors that determine the value of the coercive force, the first thing to note is that the highest value can only be reached in a perfect crystal. Lattice defects can reduce the wall energy very considerably. Bloch walls are then easily created and strong wall-pinning can also occur. The hysteresis loop will then look like that shown in *fig. 13* (compare fig. 11). It may also be found, however, that the defects have little effect on the wall energy. In this case the nucleation field-strength H_n is high and pinning will often not be observable because the pinning field H_p is weaker in absolute terms than H_n (*fig. 14*). The size of the crystal is irrelevant here. Nevertheless it is found in practice that small-grained material gives the best

magnetic properties. This may be connected with the fact that imperfections are more likely to be present in a larger crystal.

A different criterion was formerly used for determining whether a material was likely to have a high coercivity. It is possible to deduce from a comparison of energies whether or not a wall can exist in a crystal [22]. The result shows that walls can only occur in crystals with diameters larger than a certain critical value. Below this critical diameter there could only be uniform rotation of the magnetization, at $H_c = H_a$. The predictions based on this reasoning often turned out to be wrong. The mistake was that only energies were compared. The question to be asked, however, is what *force* has to be exerted to initiate a certain process (creation or displacement of a wall). The *work* performed by this force is not so relevant, although of course the process must be energetically possible. The force is closely related to the nature and number of the defects present.

Homogeneous wall-pinning

To conclude this section on crystal anisotropy something should be said about an effect observed in extremely thin domain walls. In most cases the thickness of a wall is large compared with the distance between the atomic carriers of the magnetic moments; in a perfect crystal a wall can then move without friction and is not pinned. If however the energy that the crystal anisotropy represents is high enough to be comparable with the energy of the ferromagnetic coupling between the atomic moments, then wall thickness and interatomic spacing will have the same order of magnitude. In this situation wall-pinning is also likely in perfect crystals, since the wall has to overcome an energy barrier every time it travels over one interatomic distance. We thus have a situation of 'homogeneous pinning', which gives rise to an intrinsic coercive force. This effect, predicted in 1970 [23], was confirmed experimentally a year later, although at low temperature [24] [25]. No theoretical arguments are known, however, that might suggest that the mechanism described could not give rise to a coercive force at room temperature.

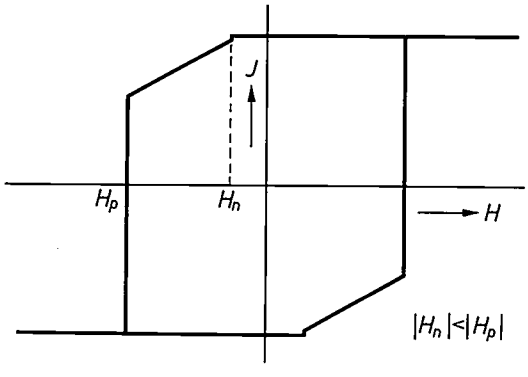


Fig. 13. Schematic hysteresis loop in which the coercive force is determined by wall-pinning. At H_n a wall is created; H_p is the pinning field.

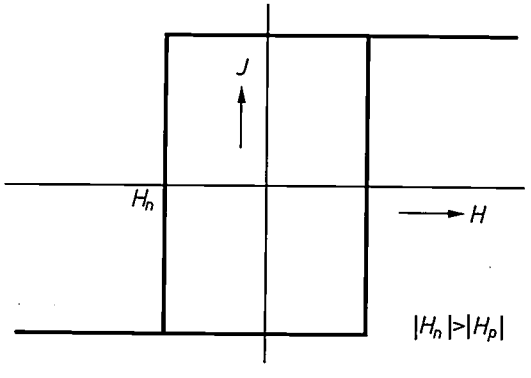


Fig. 14. Schematic hysteresis loop where the coercive force is determined by wall creation. The nucleation field-strength H_n is greater (in absolute terms) than the pinning field-strength.

Stress anisotropy

A magnetized body will in general exhibit an elastic deformation (magnetostriction) that depends on the direction in which it is magnetized. Conversely a state of elastic stress can affect the direction of magnetization, and hence the existence of stress anisotropy. This anisotropy is also formally described by an anisotropy field H_a. In a crystal with isotropic magnetostriction the relative change in length due to magnetization is given by:

$$\delta l/l = \tfrac{1}{2}\lambda(3\cos^2\Theta - 1),$$

where Θ is the angle between the direction of magnetization and the axis along which the relative change in length is measured. The factor λ is the magnetostriction constant.

It follows from the foregoing that, in the presence of a uniform elastic tensile stress σ in the material, the work E per unit volume required to rotate the direction of magnetization through the same angle is given by:

$$E = \tfrac{1}{2}\lambda\sigma(3\cos^2\Theta - 1).$$

At small values of the angle Θ this work corresponds to the work which would have to be performed by the magnetization in an effective anisotropy field whose strength is represented by:

$$H_a = 3\lambda\sigma/J_s.$$

Magnetostriction is a weak effect; λ is of the order of only 10^{-4}. This cannot be compensated by increasing

[22] C. Kittel, Rev. mod. Phys. **21**, 541, 1949.
[23] H. Zijlstra, IEEE Trans. **MAG-6**, 179, 1970, and also J. J. van den Broek and H. Zijlstra, IEEE Trans. **MAG-7**, 226, 1971.
[24] B. Barbara, C. Bécle, R. Lemaire and D. Paccard, J. Physique **32**, colloque No. 1, 299, 1971.
[25] T. Egami and C. D. Graham, Jr., J. appl. Phys. **42**, 1299, 1971.

σ, since there are of course limits to the elastic stresses to which a material can be subjected. Even in materials that have a high yield point, such as hardened carbon steel (yield stress 10^9 N m^{-2}, polarization J_s about 2 Wb m^{-2}) the maximum possible value of H_a is only of the order of 10^5 A m^{-1} (about 10^3 Oe). This makes it unlikely that stress anisotropy will ever provide sufficient magnetic hardness for permanent magnets [26]. Higher yield points may be expected in small particles without dislocations, which could in principle lead to applications. There does not seem much prospect of it at the moment, however.

Exchange anisotropy

Finally, we shall look at exchange anisotropy, i.e. the form of anisotropy connected with the exchange interaction between two neighbouring electrons. The effect of this quantum-mechanical coupling mechanism depends on the structure of the crystal, in other words on the spacing of the atoms (ions) and on their environment. Under certain conditions the interaction may lead to the parallel alignment of the magnetic moments of the atoms (ferromagnetism). Under other conditions of spacing and environment the same kind of interaction may give rise to a magnetic ordering in which neighbouring moments are aligned antiparallel. A magnetic material in which this situation is found can be described in terms of two sublattices, each consisting of a set of identically oriented moments. The resultant polarization vectors do not necessarily have to compensate one another (ferrimagnetism): ferroxdure is in fact an example of this [2]. If compensation does occur (antiferromagnetism), a magnet of the material will then have no macroscopic magnetic moment. This does not mean that the atomic moments will have no preference for a particular direction; magnetic anisotropy is still possible in an antiferromagnet.

Suppose now that a ferromagnetic and an antiferromagnetic body are in contact with one another in such a way that exchange interaction takes place through the interface; this could for example be the case when the two crystal lattices fit each other exactly. Even if the ferromagnet itself does not have magnetic anisotropy, there may still be a directional effect on its magnetization; this is then brought about by the exchange interaction, from the anisotropy present in the antiferromagnetic body. A net magnetic anisotropy of the ferromagnet-antiferromagnet system is then measured, which is referred to as the exchange anisotropy [27].

In some cases this exchange anisotropy gives asymmetric hysteresis loops. If on rotating the magnetization the maximum torque that the ferromagnet exerts through the interface upon the antiferromagnet is not sufficient to rotate it, the system will 'remember' its original direction of magnetization and give preference to it. In such a case the hysteresis loop shows a typical shift along the field axis: the material is more easily magnetized in a particular direction than in the opposite direction. This direction is 'frozen in' when the system is cooled in an external magnetic field from such a temperature that the Néel point of the antiferromagnet is passed (the temperature above which the antiferromagnetic ordering is destroyed). At that moment the sublattices coupled to the ferromagnet choose their direction of magnetization, and they then continue to determine and maintain the direction of easy magnetization. (It is assumed that the Curie temperature of the ferromagnet is higher than the Néel temperature of the antiferromagnet.) In extreme cases the material has only one remanence. The points at which the hysteresis loop cuts the field axis will then both lie to the left of the magnetization axis. The anisotropy in such a case is referred to as unidirectional.

The effect was first observed in a powder of surface-oxidized particles of cobalt [27]. The antiferromagnetic CoO skin was found to contribute to the magnetic anisotropy of the ferromagnetic cobalt grain (*fig. 15*).

This discovery was followed by many publications describing the same effect in other systems. The displaced hysteresis loop is usually put forward as evidence of the effect. However, a displaced loop may also be found when the strength of the external field is insufficient. In this case, what is measured is not the real hysteresis loop but an inner loop, which is generally displaced with respect both to the field axis and the axis of magnetization. This accounts for several of the results described in the publications mentioned.

Materials giving exchange anisotropy have not found any technical applications as magnets. They have the disadvantage that part of the magnet volume is taken up by material that does not contribute to the magnetic moment. Because of this the energy product remains fairly low. Another disadvantage is that metal/metal-oxide systems are not stable. The originally sharp interface becomes blurred by the formation of intermediate oxides, which interfere with the mechanism.

This brings us to the end of the discussion of the various forms of anisotropy which, in broad lines at least, provide a good explanation of the permanent magnetism of existing materials. Nevertheless, much work still remains to be done, including fundamental research, if we are to be able to predict exactly what the hysteresis properties of a particular material will be. Such studies will always have to be based on the structural properties of the material on a microscale.

Some fundamental questions that will have to be settled by future resarch have been touched on in the

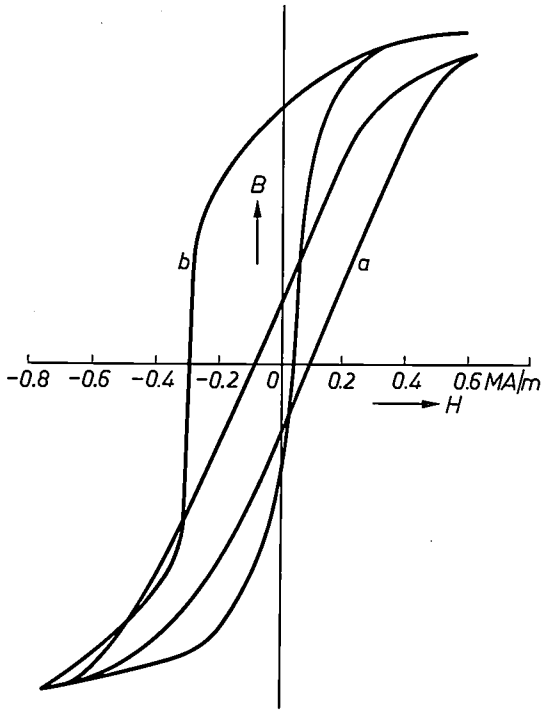


Fig. 15. Hysteresis loops of powders consisting of surface-oxidized cobalt grains. Curve *a*: cooled in the absence of a field. Curve *b*: cooled in a field. The shift occurring in this case along the field axis is connected with exchange anisotropy. The powder contains antiferromagnetic CoO and ferromagnetic Co. The antiferromagnetic part of the system causes the anisotropy.

foregoing. We should like to know, for example, why a magnet that derives its properties from crystal anisotropy is preferably made from a dispersion of small particles or grains, which may subsequently have to be densified. Other questions relate to the influence of surfaces and interfaces on the creation and movement of domain walls, and to the nature of the interaction between the domain walls and crystal defects. Finally it would be useful to know the reason for the poor chemical stability of some systems that would otherwise be of interest for magnetic applications.

Obtaining answers to these questions might continue the upward trend of the lines in fig. 2 and fig. 4, and also lead to an improvement of many applications of great topical interest, such as better magnetic tapes for recording equipment and the use of magnetic bubbles as memory devices.

Perhaps in the rather more distant future we may yet be faced with the question of whether there is a sufficient supply of the raw materials that are required for making magnets of greater energy density.

Summary. Permanent magnets can convert electrical energy into mechanical energy and vice versa, exert ponderomotive forces, and align themselves in relation to a field. The suitability of hard magnetic materials for specific applications is judged in terms of the maximum energy product $(BH)_{max}$ and the maximum change (reversible and isothermal) in the magnetic free energy. The improvement of both of these figures of merit is of great technical importance. Magnetic hardness is connected with magnetic anisotropy. A discussion is devoted to present-day knowledge of the different types of anisotropy: shape, crystal, stress and exchange anisotropy. The creation and pinning of domain walls has been studied with the aid of hysteresis loops, measured on individual microparticles (about 5 μm). Further development of permanent-magnet materials will require a better understanding of the interaction of domain walls with internal lattice defects and with the surface.

[26] A case has been known for some time in which a greater value is found for the magnetostriction constant. At the 17th Annual Conference on Magnetism and Magnetic Materials in Chicago, November 1971, A. E. Clark and H. S. Belson reported a magnetostriction constant of 2×10^{-3} in $TbFe_2$ (AIP Conf. Proc. **5**, Part 2, 1498, 1972). At this value stress anisotropy could well be interesting.

[27] W. H. Meiklejohn and C. P. Bean, Phys. Rev. **105**, 904, 1957.

Philips tech. Rev. 34, 208-224, 1974, No. 8

The optical sensors of the Netherlands astronomical satellite (ANS)

I. The sun sensors

II. The horizon sensor

III. The star sensor

The attitude-control equipment of the ANS satellite — now in orbit after its successful launch on August 30th, with all its instruments in faultless operation — includes five optical sensors. One of their functions is to detect deviations of the satellite from the attitude in which the solar panels are most favourably oriented for the generation of energy. The sensors also determine the attitude of the satellite to enable its instruments to be pointed for sufficiently long periods at the objects to be observed.

The sensors are of extremely low weight and small size; these limitations are a consequence of the choice of the launch vehicle. One of the sun sensors is a small Cassegrain telescope about the size of a matchbox. The miniaturization has not prevented a high degree of accuracy from being achieved: the star sensor, for example, determines a direction with an error of less than 20 seconds of arc. In the horizon sensor all rotational movements take place in vacuum; this is novel for this type of sensor. All sensors bear the NASA designation 'space qualified'.

I. The sun sensors

A. J. Smets

When the Netherlands astronomical satellite (ANS) disengages from its launch rocket and its initial rapid rotation has almost ceased, the satellite begins to seek the Sun, i.e. it begins to move into an attitude in which the plane of the deployed solar panels (x,y-plane) is perpendicular to the Sun's rays. This sun-acquisition manoeuvre takes place in a number of phases, involving two kinds of sun sensors — the 'coarse' and the 'intermediate' sensors. The satellite is also equipped with a third set of sensors — the 'fine' sensors. The purpose of the fine sensors is primarily to hold the z-axis of the satellite *accurately* directed at the Sun; this is necessary to enable the astronomical instruments to be pointed at celestial objects [1]. A short description of the construction and operation of these three types of sun sensors is given below.

The coarse sun sensors

Each of the six faces of the satellite carries a sun sensor with a conical field of view subtending about 20% of 4π. The six sensors therefore together view the whole 4π with some overlapping.

Ir A. J. Smets is with Philips Research Laboratories, Eindhoven.

The detector for these coarse sun sensors is a chip of single-crystal N-on-P silicon [2]. A cross-section is shown in *fig. 1*. *Fig. 2* shows how the photovoltage across the PN junction depends on the angle of incidence of the light.

A level-detector circuit in the attitude-control logic circuits converts the photovoltage into a digital signal. The trigger level for this circuit is 40 mV, so that light from any direction within an angle of incidence of 60°

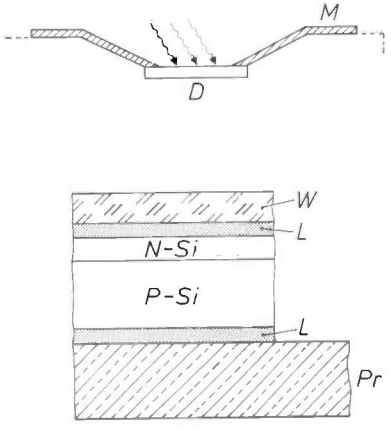


Fig. 1. *Above:* cross-section of one of the coarse sun sensors of the Netherlands astronomical satellite (ANS). *D* light-sensitive detector. *M* cone-shaped metal mask, limiting the field of view. *Below:* cross-section of the light detector, an N-on-P type silicon diode. *Pr* substrate. *L* adhesive. *W* window, made from glass containing cerium oxide to give a strong UV absorption for the protection of the adhesive. The window has an antireflective coating of MgF_2.

triggers the circuit. Any light reflected from Earth (albedo) is insufficient to trigger the circuit.

In designing photovoltaic cells for use in space applications particular attention must be paid to vibration-free mounting and to the stability of the photovoltage. Vibration tests showed that the packaging used was satisfactory. The variation of the photovoltage with temperature (it is impossible to avoid some heating from the Sun) is relatively small: 10% for a temperature variation from −60 °C to +90 °C. The decrease in photovoltage as a result of bombardment by electrons from space is insignificant. The total number of electrons incident on the sensor during the complete mission will be about 1.6×10^{13} per cm^2. This number will cause the photovoltage to fall by only a few per cent. The small value of the decrease is related to the very low oxygen content of the silicon. In addition, the window of the detector gives considerable protection from incident particles and also to some extent from micrometeorites. The window and the adhesive layer behind it have a transmission of almost 100% for visible light. Ultraviolet radiation is completely filtered out by the window; this is essential because the adhesive would otherwise deteriorate and become opaque.

The intermediate sun sensors

As a result of the combined activity of the six coarse sun sensors, the $+z$-axis soon comes within an angle of 30° to the Sun. The solar panels are then deployed. Next, the attitude of the satellite is adjusted under the control of two intermediate sun sensors (one for rotation about the x-axis, the other for rotation about the y-axis) to bring the z-axis within 1 degree of the required direction.

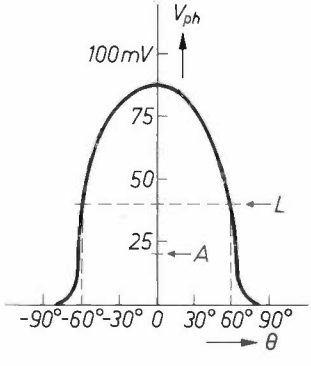


Fig. 2. The photovoltage V_{ph} generated by the photodiode of the coarse sun sensor (fig. 1) as a function of the angle of incidence θ of the light. (The voltage is derived from measurements of the current with the diode short-circuited; this is almost equal to the photocurrent at a reverse bias of up to ~500 mV.) The shape of the sides of the curve is determined by the shape of the mask (M in fig. 1). L trigger level of the detector circuit, which also affects the field of view of the detector. Earthshine light (albedo) gives signals only up to the level A (about 20 mV).

These two sensors are mounted in a metal box on the front of the satellite, that is to say the face intersected perpendicularly by the $+z$-axis. The front of the box carries two slits, one parallel to the x-axis, the other parallel to the y-axis. The sunlight incident through these slits forms linear images at the base of the box where two light-sensitive detectors are mounted. The position and motion of each of these images corresponds to the angular position of the satellite and its rate of rotation about the corresponding axis.

The satellite is in the correct attitude when each image lies exactly beneath its corresponding slit. The detectors are dimensioned in such a way that even when the satellite is ± 36° off direction (about the x-axis or the y-axis) the image still falls on the detector.

The detectors also consist of a slice of the N-on-P silicon mentioned earlier. However a slice of unusually large size (6×3 cm) is used here. Such large slices are not easy to fabricate. Six N-type regions are locally

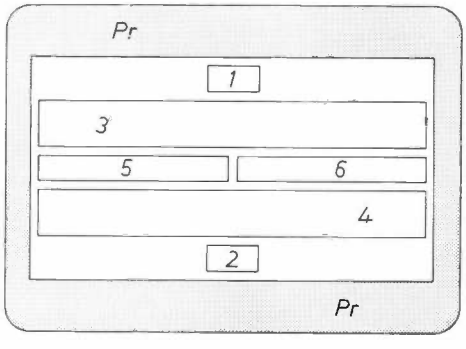


Fig. 3. The light detector for the intermediate sun sensor, showing the illuminated side of the silicon slice. Single-crystal P-type material, having a resistivity of about 10 Ω cm, is provided with six N-type regions *1-6*, each with its own contact area (not shown). Each detector thus provides six independent photovoltages. *Pr* substrate.

diffused into the slice of P-type material, each with its own contact areas (*fig. 3*). In this way each sensor consists of six independent photocells. For deriving the position and motion of the image from the succession of signals from the photocells, there is a mask with a rather intricate pattern of openings in front of the detector. This mask consists of evaporated chromium located on the lower side of the window (*fig. 4*).

The logic circuits consist mainly of twelve level-detector circuits, one for each photocell, which transform the photovoltage into digital signals. The attitude-control logic circuits examine the pattern of these sig-

[1] A general description of the attitude control of the Netherlands satellite ANS is given in: P. van Otterloo, Philips tech. Rev. **33**, 162-176, 1973 (No. 6).

[2] Photovoltaic cell, type BPX 33, manufactured by RTC, Caen, France; the surface area is reduced to 10×10 mm. The same type but of normal size is used for the solar panels.

nals to provide control signals that correct the angular deviation of the satellite from its correct orientation with respect to the x- and y-axes. The residual directional deviation of the z-axis is $\pm 0.3°$.

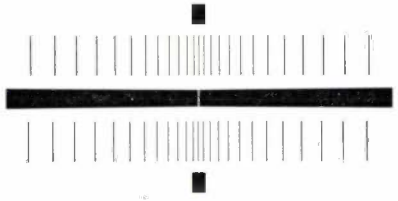


Fig. 4. Mask for the light detector in the intermediate sun sensors. The two holes at the centre correspond to the two cells *1* and *2* of fig. 3. The image of the Sun takes the form of a narrow line across the mask. The satellite has the correct attitude when this image, for both intermediate sensors, lies exactly across the two central holes and between the two cells *5* and *6*. From the rate at which the image passes over the fine slits of the mask, detected by photocells *3* and *4*, it is possible to derive the angular velocity of the satellite.

The fine sun sensors

The two fine sun sensors perform the same function as the sensors just described but sixty times more accurately [3].

The onboard computer processes the data from the fine sensors, which controls the attitude of the satellite with limit cycles restricted to a few minutes of arc. This high accuracy allows the star sensor to recognize and follow certain guide stars, as necessary in astronomical observations.

The angle of view of the y-sensor is $2°20'$, so that it is possible to detect angular displacements about the y-axis of up to $\pm 1°10'$. Such a large angular range includes many potential guide stars; this simplifies the task of the star sensor. The resolving power of the y-sensor is 42 seconds of arc. The large angular field of view and the high resolution permit the satellite to be somewhat tilted with respect to the x,y-plane ('offset' facility [1]).

The x-sensor is identical to the y-sensor, even though the resolution for rotation about the x-axis does not need to be so high.

The fine sun sensors are based on a principle not previously used for this purpose. The detector used consists of a linear array of 200 photodiodes. This array forms part of a dynamic shift register [4], each diode having a bistable flipflop circuit. The Sun produces an image of a band of light perpendicular to the array so that only a few of the diodes are illuminated (*fig. 5*). The illuminated diodes conduct, causing the flipflops to flip. The number of unchanged flipflops between the band of light and the output of the register is a measure of the position of a light/dark transition and hence of the attitude of the satellite. This number is determined by the normal shift operation of the register: it is given by the number of shift pulses necessary to shift the unchanged flipflops out of the register.

The optical system of the fine sun sensor consists of a small Cassegrain telescope having cylindrical instead of spherical mirrors (*fig. 6*). This telescope, fabricated from a single block of aluminium, is extremely compact. The spherical aberration is small and there is of course no chromatic aberration. Dimensional changes due to temperature fluctuation have no perceptible effect on the image.

Computations of the ray paths have shown that the linear image formed of a point source has a width of about 12 μm as a result of imaging errors. Diffraction effects, resulting from the small aperture of the system, which also degrade the sharpness of the image, are of the same order of magnitude.

Experimental tests have confirmed that the optical errors are small compared to the 'bit width', i.e. the length taken up by one bit of information in the detector (the distance between two diodes). The resolution is not therefore degraded by the optical errors. A temperature variation from −10 °C to +40 °C was found to cause a displacement of only 2 μm in the image.

The fine sensor has external dimensions of only $10 \times 4 \times 2$ cm. The weight is 95 grams. Its power dissi-

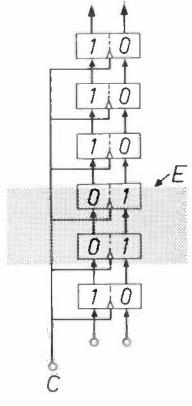


Fig. 5. Principle of the detector for the fine sun sensor. The detector consists of a linear array of 200 bistable elements (only six are shown here), which comprise a shift register. An image of the Sun in the form of a narrow band of light falls across the array. When an element is illuminated it flips over to its other state. The number of unchanged elements between the edge E of the Sun's image and the output of the register is a measure of the attitude of the satellite. C input for pulses that transport the information through the shift register.

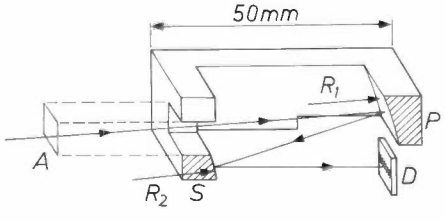


Fig. 6. Cylindrical Cassegrain telescope used in the fine sun sensors. Light from the Sun enters the telescope via the rectangular diaphragm A (7×5 mm). The field of view is about 2 degrees. P primary mirror (radius of curvature $R_1 = 123.28$ mm), S secondary mirror (radius $R_2 = 82.18$ mm). These mirrors are parts of two cylindrical surfaces having a common axis of curvature. The telescope is made from a solid block of aluminium. It gives a linear image of the Sun on the detector D (cf. fig. 5). The detector surface is located a distance 0.8 mm in front of the primary mirror.

pation is 200 mW. Measurements on three models of the sensor show that all the specified requirements have been met [5]. Temperatures from −10 °C to +40 °C are permissible and the sensor will not be damaged by shock during the launch.

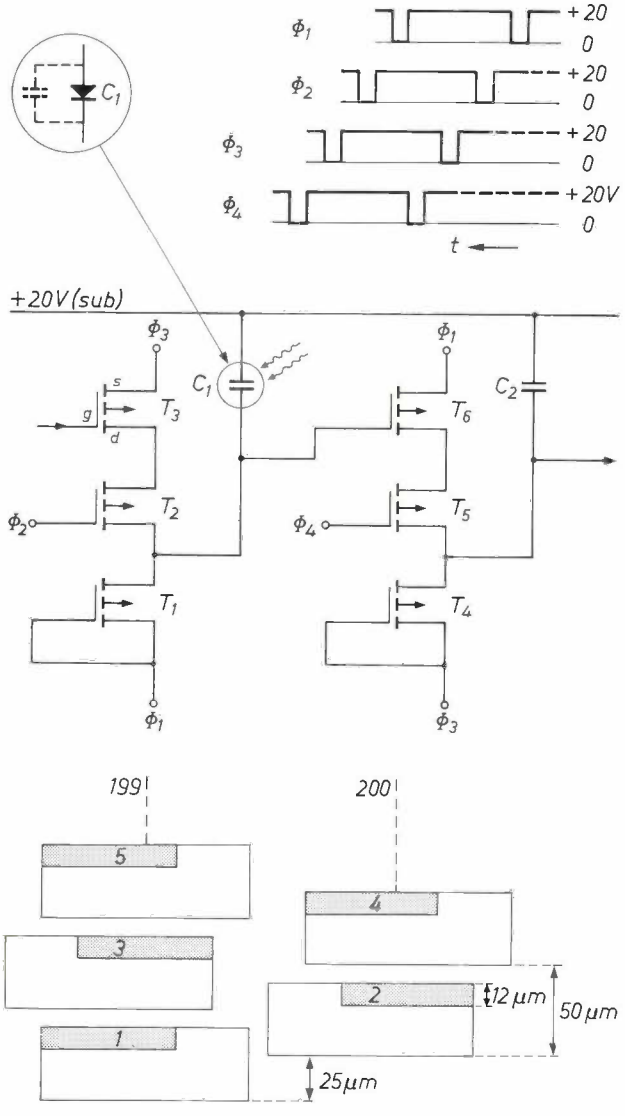


Fig. 7. Layout and operation of the detector for the fine sun sensor. The shift register consists of two parallel arrays of 100 cells, with the cells of each array displaced by 25 μm with respect to those of the other. The numbered grey areas are the photosensitive regions, one per cell. The edge of the Sun's image, i.e. the light/dark transition, falls across the cells of both arrays. The resolution of the detector is therefore 25 μm, half the effective length of one cell. This smallest detectable displacement of the image over the register corresponds to a rotation of the satellite about the relevant axis of 42 seconds of arc. The diagram shows the circuit of one cell and two repetitions of the clock pulse. The circuit consists of two almost identical parts, each with a memory element: the photosensitive capacitor C_1 (a photodiode) and the capacitor C_2. The transistor switches T_1, T_2, T_3 and T_4, T_5, T_6 together with the control pulses Φ_1, Φ_2 and Φ_3, Φ_4, ensure that the information content appears in the two capacitors (charged corresponds to dark, uncharged to light) and also shift the contents to the following cell in the register. For both registers the input to the first cell is permanently fixed at 20 V. The transistors are enhancement-type MOS transistors with a *P*-channel (in the absence of a signal these transistor switches are open-circuited). *s* source. *d* drain. *g* gate. *sub* substrate.

Design and operation of the shift register

Fig. 7 shows the circuit of the shift register and its layout, and how it operates. As can be seen, the 200 cells are not arranged in one line but in two lines of 100, with the cells of each line displaced a half-period with respect to each other. This was necessary to achieve the specified resolution while keeping the total length of the register to within 5 mm — twice this length would have led to technical difficulties.

The register is arranged as a four-stage register, i.e. a displacement of information from one cell to the next requires a group of four pulses — corresponding to four clock pulses — each pulse being applied simultaneously to all the cells. This displacement takes place in the two arrays independently, so that the whole information content shifts through the register after 100 clock pulses, i.e. after about 6 ms. To prevent the bit from cell 200 from appearing at the output simultaneously with that from cell 199, the 'odd' array is given an extra delay (a 'half cell' is added). The output signal thus consists of a series of pulses corresponding

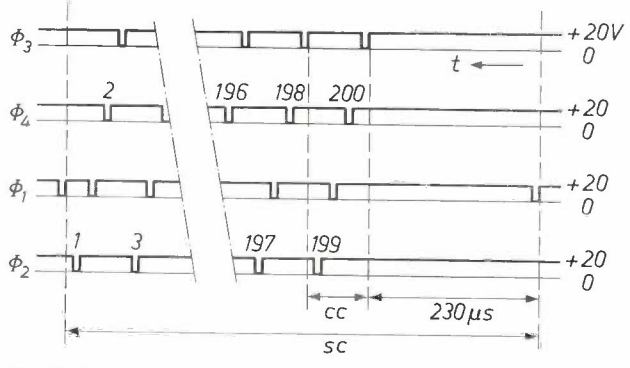


Fig. 8. A complete sequence (*sc*) of shift pulses. During the passage of the sequence, the whole of the information in the shift register (fig. 7) moves through the register to the output. The sequence begins with a Φ_1 pulse that charges all the C_1 capacitors. Next there is a period of 230 μs (4 clock pulses) during which the capacitors C_1 illuminated by the Sun are discharged. Then 100 clock pulses (*cc*) follow, so that the pulses appear in the sequence Φ_3, Φ_4, Φ_1, Φ_2, Φ_3, Φ_4, Φ_1 . . . , etc. The numbers 200, 199, . . . , 1 define the instant t at which information appears at the output and from which cell it originates.

[3] This high accuracy is necessary because the observation direction of the astronomical instruments (the *x*-axis) is coupled rigidly to the *z*-axis of the satellite. 'Rigid' in this context implies that the tolerance is at most 0.01 degree. This is achieved by attaching the sensors to one of the supports for the astronomical telescope of the satellite [1].

[4] In a dynamic shift register the information has to be continually renewed and reapplied; otherwise the information simply passes through and out of the register and is lost. On the other hand, shift registers have several advantages, e.g. a low energy dissipation per cell and a high component density, of particular importance for space applications. See, for example, L. M. van der Steen, Digital integrated circuits with MOS transistors, Philips tech. Rev. **31**, 277-285, 1970, in particular p. 281 *et seq.*.

[5] The prototype of this detector was designed by Ir H. Heyns and A. van Dijk of Philips Research Laboratories, Eindhoven. The flight model was constructed and tested by the Solid State Special Purposes Group of Philips Elcoma Division, Nijmegen.

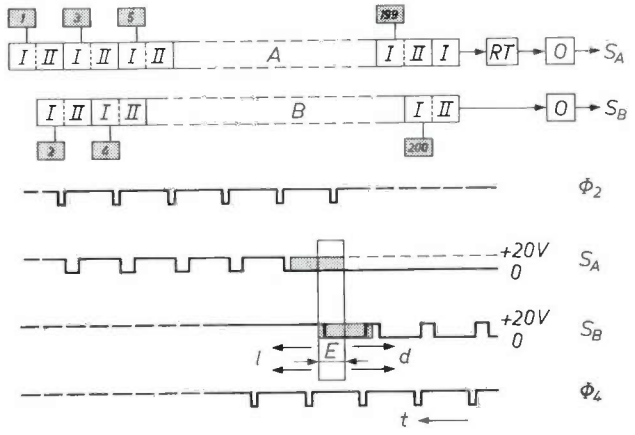


Fig. 9. An example of the voltage waveforms S_A and S_B appearing at the output of the shift registers *A* and *B*. *E* is the width of the zone within which the edge of the Sun's image can be localized. The signal in the region *d* comes from the cells on which no sunlight falls; the signal in the region *l* comes from illuminated cells. The 200 cells each consist of two almost identical parts, *I* with the photosensitive memory element C_1, and *II* with the memory element C_2 (see fig. 7). The register *A* is extended by half a cell to give the required delay between S_A and S_B (see text). The effective signal inversion thus occurring is cancelled by the inversion network *RT* (real-time inverter). *O* low-resistance output stage (with inversion). The shift register is operated by four control signals; of these only Φ_2 and Φ_4 are shown, the two signals that make it possible to detect the light/dark transition in S_A and S_B. Because of the double inversion, the signal S_A is equal to the voltage on the lower plate of the capacitor C_1 of the (delay) half cell. In the dark portion of the register all the C_1 capacitors are continuously charged; while this information is being shifted, S_A remains at 0 volts. When the information from the illuminated portion reaches the half cell, its capacitor C_1 will be discharged by the pulse of Φ_2: the signal level of S_A then becomes 20 volts. The signal S_B is the inverse of the voltage on the lower plate of the capacitor C_2 in the last cell of shift register *B*. While the information of the dark portion is still arriving at the output, Φ_4 continues to discharge the capacitor C_2, after the previous charging by Φ_3. When the information of the illuminated portion arrives at the output, Φ_4 can no longer discharge the capacitor, so that S_B remains constant at 20 V.

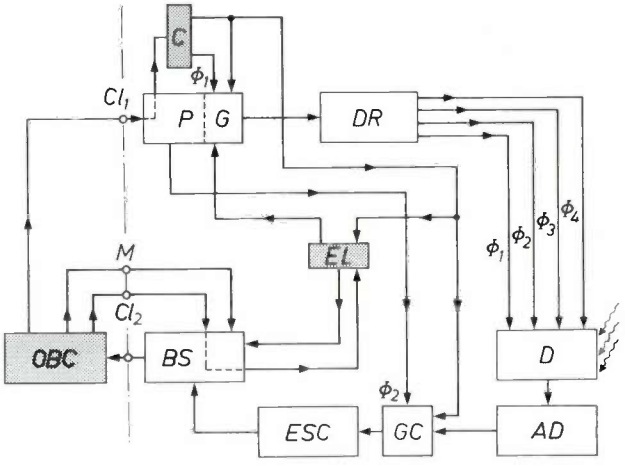


Fig. 10. Simplified block diagram of one of the fine sun sensors (to the right of the dotted line) showing how it is connected to the onboard computer (*OBC*). The electronics of the sensor includes five transistors, eleven resistors and fourteen low-power integrated circuits. The shaded blocks represent the control circuits. *D* light detector of the sensor. *ESC* sun-edge counter, capacity 8 bits; reset takes place when 200 pulses have been counted (no sun edge on detector) or at the beginning of a new shift sequence, after the information about the position of the previously established edge of the Sun has been transferred to the buffer memory *BS*. This memory is a shift register with a capacity of eight bits in series. Data is fed in in parallel and the output is in series.

The onboard computer controls the whole sensor by means of two clock-pulse trains. The clock-pulse train Cl_1 (repetition rate 65.536 kHz) determines the timing of the four-phase clock-pulse sequencies which shift the information through the register. The clock-pulse train Cl_2 (524.288 kHz) transfers the information from the buffer memory to the computer once per second (this process takes 32 μs, during which the operational status M = '0' occurs). In addition the control includes the sequence counter *C* and support logic circuit *EL*.

The sequence counter *C* counts 4 + 100 clock-pulse sequences (forming one shift sequence) and thus controls the whole internal timing of the sensor. The logic circuits are responsible for the first charging pulse Φ_1 at the beginning of the shift sequence and the reset (M = '1') and the reading (M = '0') of the information stored in the buffer memory. The computer determines which value of M is applicable.

The other five blocks represent pulse circuits. *P* pulse generator for the four-phase clock pulse trains (at 5 V level). *G* gate controlled by the sequence counter and the logic circuits. *DR* driver stage giving the clock pulses (20 V) sufficient power to provide the saturation current in the MOS transistors of the photodetector. *AD* matching circuit to bring the voltage level back to 5 V. *GC* gate controlled by the sequence counter. Application of the signal Φ_2 changes the output of the detector in such a way that the determination of the position of an edge of the Sun's image is reduced to a simple counting of pulses. The number of pulses counted is equal to the number of non-illuminated photodiodes (bit-positions) in front of the image of the Sun on the detector.

to the contents of the cells in the correct sequence (200, 199, . . . , 2, 1).

The detection operation proceeds as follows. First a pulse Φ_1 is applied which charges all the capacitors C_1 (fig. 7). No further pulses arrive during the next 230 μs (= 4 clock-pulse periods) and the capacitors C_1 of the illuminated cells are discharged. The 100 clock pulses are now applied in the sequence Φ_3, Φ_4, Φ_1, Φ_2, Φ_3, Φ_4, Φ_1, . . . etc. A complete sequence of these shift pulses is shown in *fig. 8*. Examples of the pulse trains appearing at the output of each of the 100-cell arrays are shown in *fig. 9*; the output signal is the OR output of these pulse trains. Finally *fig. 10* shows a simplified block diagram of the electronics of the fine sun sensors.

The effect of one complete clock-pulse sequence on a cell is summarized in the Table below. The information initially present at the input appears at the output at the end of the sequence. Half-way it is present in the capacitor C_1, in which however, '1' and '0' are interchanged. The *P*-channel MOS transistors conduct only when the voltage on the gate electrode is sufficiently negative with respect to the source electrode. It can be seen that the pulses Φ_1 and Φ_3 *always* charge the capacitors C_1 and C_2 respectively, whatever the input voltage of the 'lower' plate of C_1; in fact at least one of the two series switches T_2, T_3 or T_5, T_6 is opencircuited (T_2 or T_5). When Φ_2 and Φ_4 close these switches, C_1 and C_2 can again discharge, assuming that T_3 and T_6 conduct.

control pulse	V_i volts		C_1	C_2	V_o volts	
Φ_1	20	'0'	'1'	–	–	–
	0	'1'	'1'	–	–	–
Φ_2	20	'0'	'1'	–	–	–
	0	'1'	'0'	–	–	–
Φ_3	20	'0'	'1'	'1'	0	'1'
	0	'1'	'0'	'1'	0	'1'
Φ_4	20	'0'	'1'	'0'	20	'0'
	0	'1'	'0'	'1'	0	'1'

II. The horizon sensor

P. van Dijk

Attitude measurement and horizon sensor

Searching for an astronomical object occurs under the control of the horizon sensor; the orientation involves rotation of the satellite about its z-axis only (this points towards the Sun). The ANS satellite can execute three such manoeuvres: the 'slew', 'scan' and 'slow-scan' operating modes [1]. The onboard computer controls these manoeuvres, use being made of the angle between the line of sight of the Cassegrain telescope (i.e. the x-axis) and the local vertical. This angle represents the attitude information from the horizon sensor; its value follows from the two angles (*fig. 1*) between the line of sight mentioned above and the Earth's horizon, which are determined once per second by the horizon sensor.

Between about 25 and 50 km above the Earth's surface, the intensity of the infrared emission from atmospheric carbon dioxide in the 14-16.5 μm band decreases rapidly almost to zero (*fig. 2*). This rapid decrease represents quite a sharp transition and forms the most stable horizon indicator known. Neither geographical position nor weather — nor even day and night — have much effect on the position of the transition [2] [3].

Fig. 3 shows a cross-section of the horizon sensor, together with its drive motor and the electronic units that provide the data relating to the measured angles. The sensor head protrudes from the dark rear wall of the satellite and observes the sky via a rotating mirror at 45°. The infrared horizon is detected twice per revolution: at the space/Earth transition (angle α, fig. 1) and at the Earth/space transition (β). Because the mirror scans the whole x,y-plane, the sensor cannot fail to detect the horizon (assuming the z-axis points towards the Sun).

The infrared optics of the horizon sensor besides the mirror are a concavo-convex germanium lens, a band-pass filter and a hemispherical germanium lens that carries the detector. The electronics includes two counters that start to count as soon as the mirror 'looks' in the $+x$-direction of the satellite. When the first horizon crossing is seen, the output signal from the detector stops the first counter. The second counter carries on counting until the next horizon crossing is observed (inverse output signal). The numbers of pulses counted give a measure of the angles α and β.

Ir P. van Dijk is with Philips Research Laboratories, Eindhoven.

The two-stage reduction gear and the ball-bearings are vital components in this mechanically scanning system. They operate in vacuum, with dry lubrication. Dry lubrication was preferable because it does not

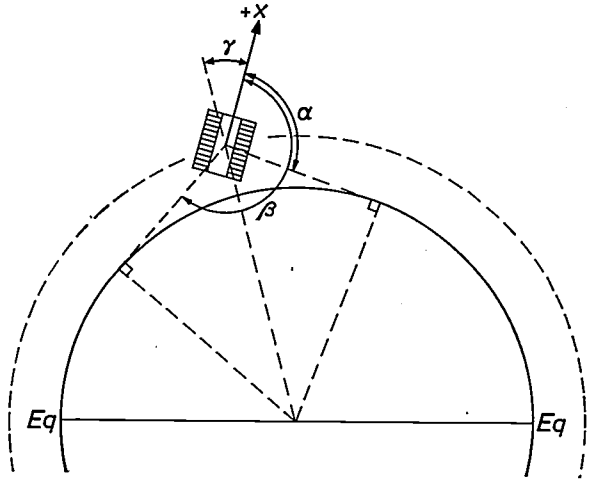


Fig. 1. Attitude control of the satellite by means of the horizon sensor. The astronomical instruments look in the direction of the arrow, the positive x-axis. The z-axis (perpendicular to the plane of the drawing) points to the Sun. The horizon sensor detects the position of the infrared horizon in two directions. From the measured angles α and β, the onboard computer calculates the angle γ between the x-axis and the local vertical. *Eq* equator.

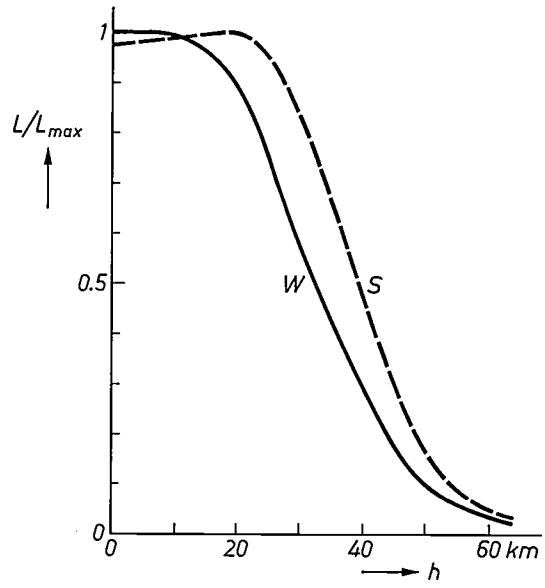


Fig. 2. The normalized radiance L/L_{max} in the wavelength range 14-16.5 μm (atmospheric carbon dioxide) as a function of the height h in km above the surface of the Earth. $L_{max} \approx 7.5$ W/m² per unit solid angle. The curves refer to the northern hemisphere, S for the summer and W for the winter. Both are mean intensity profiles, based on a large number of meteorological observations [3]. The horizon sensor detects a horizon transition (space/Earth, or Earth/space) when $L/L_{max} = 0.5$.

[1] P. van Otterloo, Attitude control for the Netherlands astronomical satellite (ANS), Philips tech. Rev. 33, 162-176, 1973 (No. 6).

[2] J. Breton, Detection of local vertical aboard space vehicles, Acta Electronica 13, 217-226, 1970. (Also in French, pp. 207-216).

[3] F. Desvignes, Rayonnement terrestre et senseurs d'horizon, Acta Electronica 13, 227-247, 1970.

contaminate the mirrors and lenses in the satellite. In the six months that the satellite has to remain operational the electric motor will make five hundred million revolutions; life tests have shown that this is no problem.

The horizon sensor will now be discussed in more detail. The most important characteristics and data relating to the horizon sensor are summarized in *Table I*. The values achieved for the mass of the sensor (1700 g) and its power dissipation (0.9 W) are satisfactory for this type of sensor.

Optics and detector

The rotating mirror is of aluminium, machined optically flat, with a reflecting layer (reflection coefficient 0.97) of evaporated aluminium having a thin coating of silicon monoxide to prevent oxidation or mechanical damage. The next component, the objective of the IR telescope, is a coated high-aperture meniscus lens of single-crystal germanium [4]. This lens does not give coma and introduces only slight spherical aberration, so that the image disc is small; this is important for the signal/noise ratio. Since the refractive index of germanium is high ($n \approx 4$), the telescope is small.

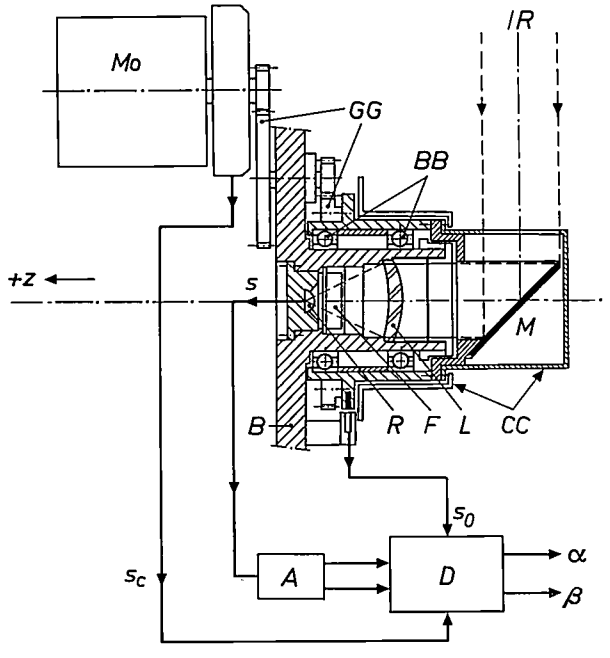


Fig. 3. The horizon sensor. The observed infrared radiation *IR* enters the instrument via the rotating mirror *M*. *CC* shield caps, gold-plated externally; these prevent excessive cooling of the sensor. *B* stationary frame. *L* single-element germanium lens. *F* bandpass filter. *R* immersion bolometer. *Mo* brushless d.c. motor. *GG* reduction gear. *BB* ball-bearings. *A* analog electronics. *D* digital electronics. At the moment when the horizon sensor looks in the same direction as the astronomical telescope (the $+x$-direction, see fig. 1), a reference signal s_0 is generated that resets the two pulse counters. The signal s_c, whose repetition rate is determined by the rate of rotation of the mirror *M*, is the input to the pulse counters. When a horizon is detected by the sensor, the bolometer signal s stops one of the counters; when the next horizon crossing occurs, the other counter is stopped. The numbers of pulses registered are measures of the two angles α and β (fig. 1).

Table I. Characteristics of the horizon sensor

Mass	1700 grams
Power dissipation (including motor)	900 mW
Permissible vibration levels	accelerations up to 16 g in band 100-200 Hz and 5 g in band 200-2000 Hz.
Permissible temperature range	−20 °C to +40 °C
Expected temperature range	−10 °C to +30 °C
Life	
Running-in (clean air) + testing	1000 hours
Mission	4400 hours (minimum)
Reliability of electronics	⩾ 99.44% for 6 months
Motor	
Power	200 mW
Torque	10×10^{-4} Nm (at lowest controlled speed)
Control range	1820-2020 rev/min
Power dissipated by electronic commutation + speed-control circuits	200 mW
Bolometer bridge	
Area of thermistor flake	0.1×0.1 mm²
Resistance of flake	250 kΩ at 25 °C
Supply voltage	2×16 V, 60% of peak voltage (thermal runaway)
Responsivity	550 V/W
Time constant	1.4 ms
Optics	
Field of view	1° × 1°
Total transmittance	50%
Power incident on detector	0.23 μW (mean)
Accuracy	
Summer-winter variation	about 4′
Alignment error	2 to 3′
Noise error (standard deviation)	about 2′
Resolution	5.3′

The bandpass filter *F* is an all-dielectric interference filter [5]. It has a transmittance of 60% for wavelengths between 14-16.5 μm and outside this passband the transmittance is less than 0.1 %. The second germanium lens and the detector form a hemispherical immersion bolometer [6]. The detector consists of a bridge circuit of two identical thermistors (*fig. 4*) in the form of thin flakes. One of these, the 'active' flake, is in optical contact with the lens. The other is mounted close to it so that the effect of the ambient temperature is the same for both flakes. The incident radiation strikes the 'active' flake only and causes a temperature rise of about 10^{-4} K. The active flake has an area of 0.1×0.1 mm² giving a field of view of about 1° × 1°. (But note that this is not the accuracy of the angle measurement, which is about 0.1°.)

The advantage of using the hemispherical optics for the bolometer is that, for a given incident power, the detector can be smaller by a factor of n^2; in the present case it is 16 times smaller. With such a reduction the signal/noise ratio is improved by a factor inversely

proportional to the square root of the area of the detector, in the present case by a factor of 4.

The supply voltage for the thermistor bridge is chosen so as to limit the dissipation in the two thermistor flakes; overheating ('burn-out') as a result of a rapid increase in current with continuously decreasing resistance is thus avoided.

The electronics

Fig. 5 is a schematic diagram of the electronics of the horizon sensor. The signal from the bolometer bridge is amplified about a hundred times in a preamplifier to about 12 mV. This preamplifier has an FET input stage giving a sufficiently high input impedance and low noise. For a source impedance (the bolometer flakes) of between 0.1 and 1 MΩ, the noise figure at the peak frequency (35 Hz) of the amplifier passband is less than 1 dB.

The supply voltages for the preamplifier and the bolometer bridge are carefully smoothed, since the signal is very small. The preamplifier has electrical and magnetic screening.

The preamplifier differentiates the signal to produce two pulses of opposite sign corresponding to the two horizon crossings (space/Earth, Earth/space). These are amplified in the next block, where an integrating network decreases the bandwidth to improve the signal/noise ratio by a factor of about two. A clipping circuit then separates the positive and negative pulses. Each pulse is then fed to its own level detector, which determines the instant at which the signal passes a given level; this is then the instant of the horizon crossing. The level is not constant; it is half of the amplitude of the previous pulse from the same horizon (normalized radiance technique). In this way variations of the amplitude caused by effects such as changes in atmospheric radiance cannot lead to gross errors of measurement. (This simple level control is possible because the amplitude never changes rapidly.) The level

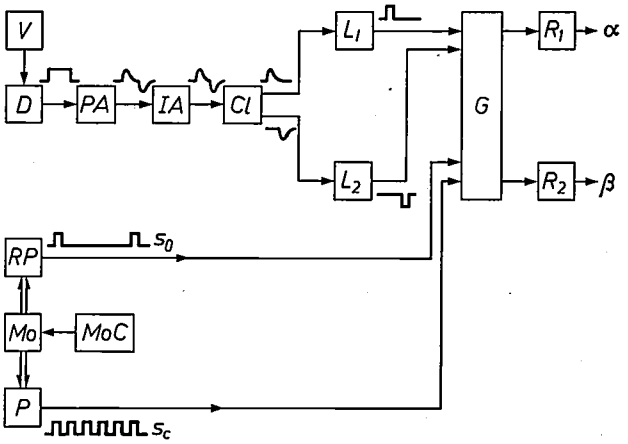


Fig. 5. Block diagram of the horizon-sensor electronics. The circuit components are mounted on five printed-circuit boards and where possible ICs are used to minimize the power consumption. *V* d.c./d.c. converter for the detector supply. *D* detector. *PA* differentiating preamplifier (fig. 4). *IA* amplifier stage with integrating action ('matched filter'). *Cl* clipping circuit. $L_{1,2}$ pulse-level detector. *G* gates. $R_{1,2}$ counters with registers for storage of the measured angles (α and β, fig. 1). *Mo* motor. *MoC* motor-speed control. *P* pulse generator, synchronous with the motor, generating the counting pulses s_c (4096 pulses per mirror revolution). *RP* pulse generator providing the reference signal s_0 (1 pulse per mirror revolution), which resets and starts the counters. The data stored in the registers is read by the computer once per second. Secondary circuits that collect and process housekeeping data (e.g. current in bolometer resistances, motor temperature) for transmission to Earth, are not shown.

detectors give pulses of width 50 μs, which stop the counters by means of gates. The contents of the counters are stored in registers and are transferred once per second to the onboard computer, for the calculation of the angle γ (fig. 1).

The pulses to be counted are generated at a rate of 4096 pulses *per revolution of the mirror*, thus ensuring that variations in motor speed have very little effect on the measurement of angle. One counted pulse corresponds to an angle of 5.3 minutes of arc. A small magnet that rotates with the mirror induces a voltage pulse in a pick-up coil whenever the mirror looks exactly in the $+x$-direction. This reference pulse resets the counters to zero and then starts them counting again.

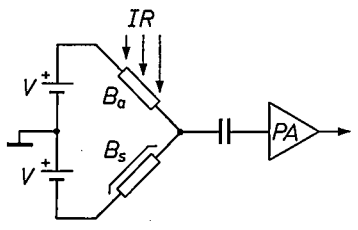


Fig. 4. The detector of the horizon sensor is a bridge circuit of two thermistor flakes (temperature coefficient about −4%). *IR* represents the infrared radiation to be detected; this causes a rise in temperature and a reduction in the resistance of the active bolometer flake B_a, so that when the horizon is crossed the voltage across the bridge changes suddenly. B_s shielded thermistor flake to compensate undesirable fluctuations. *PA* low-noise preamplifier with FET input stage. *V* voltage supply.

[4] This lens (effective diameter 20 mm, focal length 25 mm, transmittance 85%) was designed and made by J. J. Hunzinger, Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes, France.

[5] This filter (made by SEAVOM, Franconville, France) consists of more than a hundred interference layers, stacked on both sides of a germanium substrate. The filter has a transmittance of 60% for wavelengths in the range 14-16.5 μm; outside this range the transmittance is less than 0.1%. See for example P. Baumeister, Interference, and optical interference coatings, in: R. Kingslake (editor), Applied optics and optical engineering, Vol. I, pp. 285-323, Academic Press, New York 1965.

[6] See p. 353 of C. F. Gramm, Infrared equipment, in: R. Kingslake (editor), Applied optics and optical engineering, Vol. II, pp. 349-378, Academic Press, New York 1965. The immersion bolometer used was made by Barnes Engineering Company, Stamford, Conn., U.S.A.

The construction and the motor

The horizon sensor has about five hundred mechanical components. The sensor is built around a central frame (*fig. 6*); the telescope forms part of this frame and the electric motor, reduction gear and four of the six bearings are mounted upon it. The frame also carries the printed-circuit boards for the electronic circuits. The housing of the sensor does not have to be hermetically sealed; this avoids the extra complication of a window (with transmission losses) and also saves weight. Two gold-plated shield caps (fig. 3) are provided to prevent excessive temperature gradients in the sensor, which might otherwise arise because of radiation into empty space (effective temperature about 4 K).

The rotating mirror is attached to a tube rotating on two ball-bearings (1 rev/s). The tube has a gear ring attached to it, driven by a pinion on an auxiliary shaft; to obtain the speed reduction with respect to the motor shaft there is a second pinion-gearwheel combination. The ratio of the number of teeth of each pair is not an integer, to make the time between successive contacts of the same pair of teeth as large as possible and hence ensure that the wear is distributed more evenly. The motor shaft and the auxiliary shaft also run on ball-bearings.

The life to be expected from the ball-bearings and the gearwheels is more than sufficient. Materials and lubrication have been selected on the basis of tests in vacuum. In such tests the shaft of the electric motor has been run at twice normal speed for about 10^9 revolutions without any difficulties; this is twice the desired number of revolutions. During rotation there is continuous production of a solid lubricating film (transfer film) of MoS_2 and the plastic PTFE. This layer forms in the bearings as a result of the wear of the bearing cage, which is made of a composite of these two substances [7]. Such a bearing has of course to be run in for a certain length of time. The gearwheels are also fabricated of this composite material, and the pinions are made of a harder material (Cr-Ni steel). The difference in structure of these materials rules out cold welding. The choice of these materials is based on an investigation of various combinations. Continuous measurements of wear have been made electronically under simulated operating conditions [8].

Life tests have shown that the sensor can sufficiently withstand the vibration and extreme temperatures encountered in space applications.

The electric motor for driving the mirror is a brushless d.c. motor [9]. This design was selected chiefly because of its high efficiency and the absence of brush wear; it also possesses a considerable reserve of power: its maximum torque in the speed range of interest is about five times the total frictional torque due to the bearings and gearwheels. The pulse generator, mounted on the motor housing, comprises two stationary steel discs with teeth around the circumference. Two toothed rings attached to the motor shaft rotate concentrically round the discs, at a nominal speed of 32 revolutions per second. These components, together with a permanent magnet, form a magnetic circuit in which the flux varies, owing to the variation of the air gaps be-

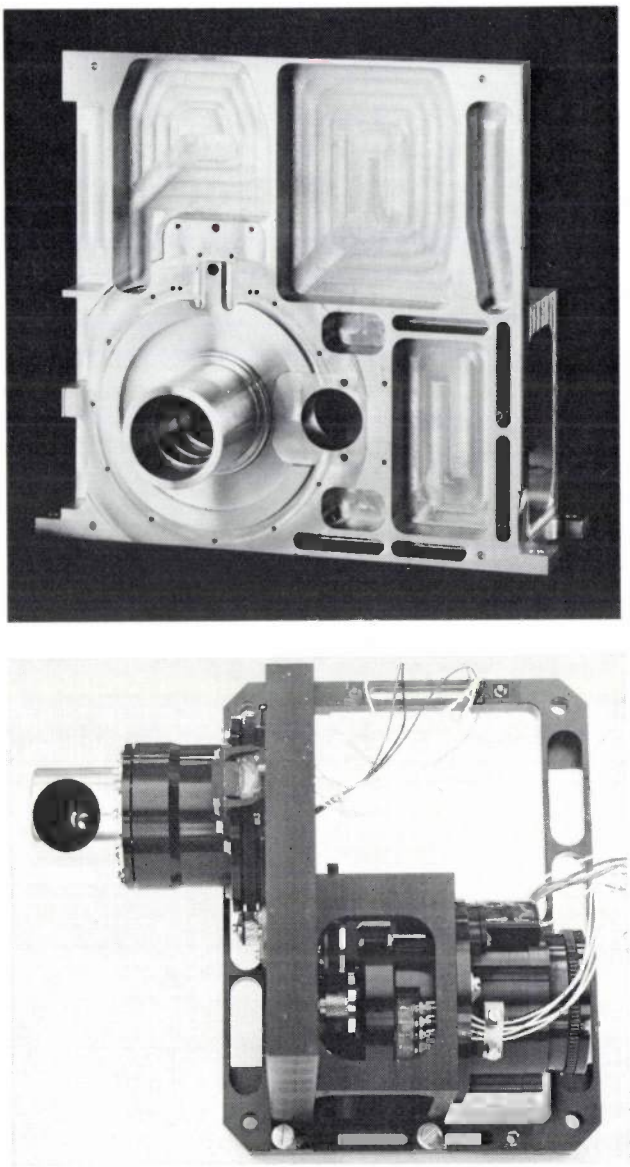


Fig. 6. *a*) Aluminium-alloy frame (14.6 × 13.4 cm), the largest component of the horizon sensor, machined by a numerically controlled milling machine. The frame is mounted on a baseplate fixed to the satellite. *b*) The frame seen from above. Upper left: rotating head with mirror. Lower right: electric motor (efficiency 80%) which drives the mirror, via a reduction gear of 32 : 1, at a speed of 1 revolution per second. The reduction gear consists of two pinion/gearwheel pairs (also visible) (184/29 and 116/23). The gearwheels have helical teeth, whose edges make contact on only one side. The torque transmission per stage is better than 90%. In the completed sensor, the space beside and above the motor is occupied by the electronics, mounted on five printed-circuit boards.

tween the teeth. Since there are 128 teeth and the rings rotate 32 times as fast as the mirror, the signal induced in an adjacent coil has 4096 pulses for each revolution of the mirror. Because of the fixed relation between the pulses from the pulse generator and the rotation of the mirror the speed of the motor requires very little regulation.

Accuracy

The resolution of a single measurement of angle by the horizon sensor is primarily determined by the quantization of the measurement: one counted pulse corresponds to 5.3 minutes of arc. The standard deviation of the error in the angle measurement due to noise in the equipment is about 2 minutes of arc.

The accuracy of the angle information is not of course entirely determined by the sensor; it also depends on the object observed, i.e. the infrared horizon. The main source of error here is the variation in height with the seasons. When, for example, the satellite passes the equator in January, the sensor views the infrared horizon in the northerly direction at its winter height, and in the southerly direction at its summer height, giving rise to a difference of up to 8′. There is then a systematic error of about 4′. In the 'scan' and 'slow scan' operating modes the systematic error only affects the *instant* of star recognition, which is fortunately not very important.

The random errors on the other hand are important, because they have a direct effect on the scanning speed of the satellite. Thus, if the scanning speed were too high, the star sensor of the satellite could easily fail to recognize the desired star pattern in the time available [1].

Finally, let us examine in somewhat more detail the contribution of instrument noise to the random error in the separate angle measurements. This error is closely related to the responsivity (i.e. the ratio of signal voltage to incident power), to the noise signal from the bolometer and to the rise time of the pulse at the input to the level detectors (fig. 5).

In discussing the detection of radiation, two quantities are commonly used:
1. The responsivity, $R = S/W$, where S is the signal amplitude in volts and W the power of the detected radiation in watts.
2. The useful sensitivity or detectivity, $D = R/N$, where N is the r.m.s. noise voltage. Instead of D, the noise equivalent power NEP is also used: $NEP = D^{-1}$, the radiation power that gives a signal voltage $S_n = N$.

The responsivity of the bolometer is 550 V/W; owing to the bridge circuit, local temperature variations have little effect. The signal amplitude has a mean value of about 120 μV, but it may drop to half of this value under conditions of poor radiance. The main contribution to the noise is from the bolometer flakes; these have an equivalent noise resistance of 125 kΩ. At a temperature of 290 °C, such a resistance gives a noise voltage of about 0.5 μV in a bandwidth of 100 Hz. At frequencies below 30 Hz the bolometer flakes exhibit some flicker noise. Owing to the measures mentioned earlier, the noise contributions from the preamplifier (FET input stage) and from the bolometer supply remain very small. All these contributions add up to a noise amplitude of less than 1 μV and the signal/noise ratio is therefore larger than 100.

The uncertainty Δt in the times of the horizon crossing, due to the noise, can be estimated from the relation

$$\tau_d/S = \Delta t/N,$$

where τ_d is the rise time of the pulse S at the input of the level detector and N is the noise voltage at the input. The error in the angle measurement corresponding to this Δt is $2\pi\tau_d N/S$ radians for a mirror rotating at 1 revolution per second. The rise time τ_d, which depends on the response time of the bolometer, the time constants in the electronic circuits, the unsharpness of the horizon (the slope of the steep part of the curve in fig. 2) and the field of view of the sensor, amounts to 4.2 ms. From the values given for τ_d and N/S we find that the angle error due to the noise should be about 1 minute of arc.

The amplitude N of the thermal noise of the bolometer flakes is proportional to $(\varrho/d)^{\frac{1}{2}}$, where ϱ is the resistivity of the material and d is the thickness of the flakes. The area of the illuminated surface of the detector has no effect on N. For most detectors, including the detector used here, the detectivity D ($= S/WN$) is found to improve as the illuminated area is decreased: D is inversely proportional to the square root of this area.

The size of the illuminated area required is determined not only by the detectivity but also by the required response time of the detector and the field of view of the sensor. For given optics, a smaller field of view — yielding a more accurate horizon measurement — is generally better, provided that the detector area is not so small that the detected power W is small compared with the noise power of the amplifier. The response time of the detector should be such that the sensor is able to follow the steep edge of the intensity curve (fig. 2) at the scanning rate used. Once the size of the bolometer flakes has been decided, there is only a restricted freedom of choice; a higher detectivity can then be achieved only at the expense of the response time and vice versa. The final choice of the detector is a compromise between high responsivity, low noise and fast response.

[7] This composite, 'Duroid 5813' of the Rogers Corp. (Rogers, Conn., U.S.A.), consists of PTFE (70%) and MoS_2 (15%), with glass fibre (15%) for reinforcement. PTFE is an abbreviation for polytetrafluoroethylene ('Teflon', DuPont), a polymer of C_2F_4.
[8] P. van Dijk, Slijtage-onderzoek aan tandwielen voor de horizonsensor van de Astronomische Nederlandse Satelliet, De Constructeur **12**, No. 6, 51-55, 1973.
[9] W. Radziwill, A highly efficient small brushless d.c. motor, Philips tech. Rev. **30**, 7-12, 1969.

III. The star sensor

W. J. Christis

Fine pointing at stellar objects

The star sensor is mounted behind the telescope whose main function is in the UV-spectroscopic investigation to be carried out with the satellite. The star sensor supplies the onboard computer with accurate information for the pointing of the telescope — the smallest measurable deviation is about 20 seconds of arc. The computer can then control the attitude of the satellite to such an accuracy that a given object remains within 1 minute of arc on the line of sight of the telescope. In order to achieve this control, a fixed direction is of course necessary as a reference. For this purpose the star sensor, during each measurement, makes use of two relatively bright guide stars in the neighbourhood of the object to be measured, since this object will usually be too faint to provide its own attitude-control signal [1].

The telescope performs its two functions with the aid of an oblique beam-folding mirror with a small central aperture (see *4*, fig. 12 in the first article of note [1]), situated in the focal plane of the telescope. The light from the observed object passes through this aperture to reach the spectroscope, which is situated on the axis of the telescope, behind the focal plane. Stars appearing in the direct neighbourhood of the object are imaged via the folding mirror on the photocathode of an off-axis image-dissector tube — the detector of the star sensor. The central aperture in the mirror gives rise to a blind spot in the centre of the field of view. The satellite is manoeuvred until the object to be observed lies exactly in the blind spot. The manoeuvres are carried out until the image of a certain neighbouring star (the tracking star) known to the computer lies at a prearranged position on the photocathode; this star image is then held in the same position for the whole measurement (up to half an hour). Two operating modes are involved: recognition and tracking.

The tracking star is identified by the recognition of the star pattern formed by the tracking star and a second star, which we call the recognition star. The criterion for identification is the imaging of two stars, within a period of 1 second, on the photocathode, at locations whose spacing corresponds to the known positions in the sky of the two stars. This means that the angular separations ϕ and θ, which are known to the computer, agree with the differences in z-coordinate and y-coordinate, respectively, of the two images (*fig. 1*).

In the tracking mode, the position of the tracking star on the photocathode is measured continuously; this is done by keeping a scan pattern centred on the image of the star.

If the onboard computer concludes that the measured position is not correct — so that the object under observation threatens to move out of the central blind spot — the computer sends control signals to the reaction wheels [2] to correct the attitude of the satellite.

Construction and operation

Fig. 2 includes a block diagram of the star sensor, showing the general principles of its operation. It can be seen that the electronic units are controlled directly by the computer; measurement data is fed back to the computer at least once per second during both recognition and tracking. (The sampling period of the complete attitude control is 1 second [1].) The image-

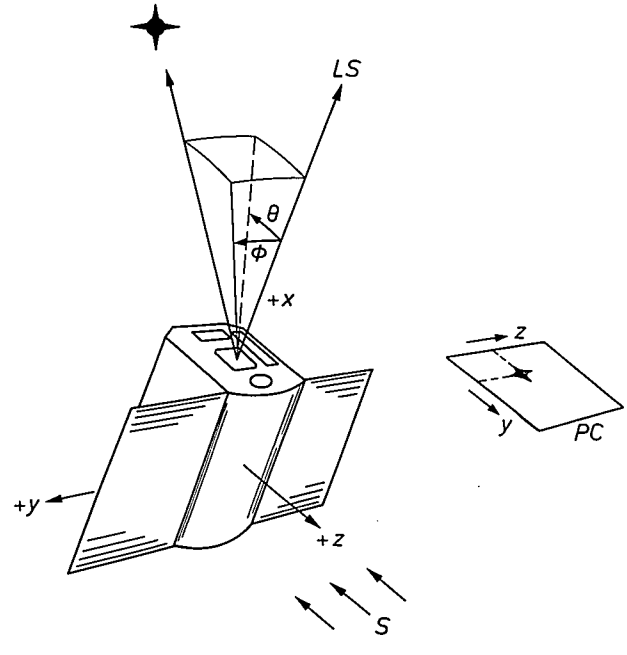


Fig. 1. The satellite ANS carries a fixed Cassegrain telescope fitted with a UV spectrometer for the investigation of young hot stars. *LS* represents the line of sight of the telescope and is coincident with the x-axis of the satellite. The z-axis of the satellite is kept pointed at the Sun (*S* represents the incident sunlight) within an error of at most 0.01 degrees. As soon as the fine pointing comes into operation, *LS* will be fixed to a stellar object to an accuracy of 1 minute of arc. This is done with the aid of data from the star sensor, which also makes use of the telescope. The smallest perceptible direction error corresponds to 1 mm at a distance of 10 m ($\approx 20''$). A star has angular coordinates θ and ϕ with respect to *LS*; the corresponding image on the photocathode *PC* has coordinates y and z. The coordinate y thus corresponds to the angular coordinate θ, which is a measure of the rotation of the satellite about its y-axis; similarly, ϕ corresponds to z and to a rotation about the z-axis.

Ir W. J. Christis is with Philips Research Laboratories, Eindhoven.

dissector camera tube [3] contains, in addition to the photocathode, an accelerating electrode, a drift tube with deflection coils, a diaphragm with a central aperture and an electron multiplier. The electrons liberated by light are accelerated towards an electrode at positive potential with respect to the photocathode. Because of the focusing action of the coils and the deflection of the beam — which is adjustable — only those electrons originating from one small region of the photocathode can pass through the aperture in the diaphragm. These electrons then enter the electron multiplier where secondary emission gives an effective amplification of about one million, yielding a video signal at the output. All other electrons are lost. By adjustment of the current in the deflection coils the location of the small region on the photocathode that gives rise to the video signal can be varied at will. The complete photocathode can thus be scanned. There is no storage facility in this type of camera tube; ghost images (images from locations not corresponding to the present telescope attitude) therefore do not occur. The photocathode has such a fast response that any arbitrary position can be scanned at any moment. In the recognition mode the scan follows two short parallel lines that are crossed by the star images at right angles; tracking is done with a cross-shaped scan pattern that locks on to the tracking star.

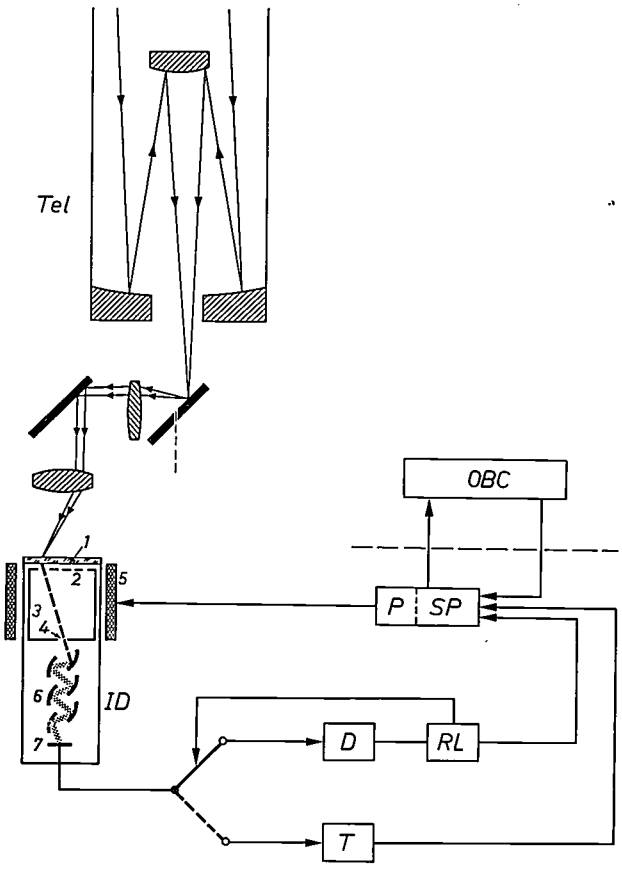


Fig. 2. Block diagram of the star sensor with its controlling system, the onboard computer *OBC*. *Tel* telescope. *ID* image-dissector camera tube with *1* photocathode, *2* accelerator electrode, *3* drift tube, *4* diaphragm with aperture, *5* deflection and focusing coils, *6* dynodes, *7* anode. The dynodes and the anode form the electron-multiplier section of the tube. The power supply for the focusing and the high voltage for the dynodes are not drawn. *P* power supplies to the deflection coils of the image-dissector tube; *SP* the pattern generators for scanning the photocathode; *P* and *SP* are in duplicate (one for the *y*- and one for the *z*-coordinate). *RL* recognition logic, controlling the recognition of a tracking star (recognition mode) and taking the subsequent decision to commence tracking this star (tracking mode). *D* detector circuits that detect the passage of a star over a scan pattern on the photocathode during the recognition mode. *T* tracking circuits that centre the scan pattern on the image of the tracking star in the tracking mode. The onboard computer thus regularly obtains coordinate data relating to the tracking star.

Apart from the input and output circuits for the onboard computer, the electronics of the star sensor can be divided into three parts: the power supplies and their control circuits, the detection circuits and the recognition logic, and the tracking circuits. In the power-supply circuits, the supplies for the deflection coils with the scan-pattern generator are of special importance.

The detection circuits comprise a video amplifier with a preamplifier, an integrating filter and a threshold detector. The gain of the video amplifier can be increased by the computer during the mission. Such an increase may be necessary if there is any decrease in the sensitivity of the camera tube. The recognition logic opens and closes switches to cause the recognition operations to take place with the correct timing, and to start the subsequent tracking mode. The amplifiers mentioned above are also used in the tracking mode. The other tracking circuits include two compensating networks — one for the *y*- and one for the *z*-coordinate — to shape the response in such a way that the remaining deviations between the image on the photocathode and the scan pattern that tries to follow the image are reduced sufficiently rapidly to zero. The output signals from these two tracking filters control the scan-pattern generator; two level detectors decide whether the image on the photocathode is displaced by a distance greater than one resolution element, i.e. whether it has passed more than one scan position.

[1] A general description of the attitude-control system of the ANS satellite is given in: P. van Otterloo, Philips tech. Rev. 33, 162-176, 1973 (No. 6).
A more detailed description of the telescope is given in: J. W. G. Aalders, R. J. van Duinen and P. R. Wesselius, The Groningen ultraviolet experiment with the Netherlands astronomical satellite (ANS), Philips tech. Rev. 34, 33-42, 1974 (No. 2/3).

[2] The reaction wheels are described in: J. Crucq, Philips tech. Rev. 34, 106-111, 1974 (No. 4).

[3] Manufactured by ITT (Fort Wayne, Ind., U.S.A.), type F4012, provided with magnetic focusing and deflection. The scanned area of the photocathode (S20) is 10.8×10.8 mm^2; in the present case 256×256 separate scan positions can be selected. The tube has the sensitivity of a photomultiplier and an imaging performance comparable with that of other good camera tubes.

Every such step is registered by the y counter or the z counter. At the same time the data from these counters is fed to digital analog converters in which the values (the changes in the coordinates of the tracking star on the photocathode) are transformed into a correction signal in the deflection coils. The scan pattern can thus remain centred on the tracking star even when it is displaced over more than one resolution element. The contents of the y and z counters are transferred to the onboard computer, which uses the coordinate data to calculate the attitude-control signals for the reaction wheels.

The scan field on the photocathode of the camera tube comprises 256 steps in both directions; one step is of course equal to one resolution element. There are thus in total 2^{16} possible positions in the scan pattern; the selection of any such position takes place by means of the y and z counters. The state of these counters thus represents the Cartesian coordinates of the corresponding position on the cathode. The direction of scanning, the scan speed (256 positions per second) and the step size are determined by the scan-pattern generator. The image of a star on the photocathode is almost circular and has a diameter nominally equal to two resolution elements. The central aperture in the diaphragm behind the drift tube is of such a size that the electrons passing through it have all originated from a circular patch of the photocathode of diameter equal to 8 times the resolution element. In a rapid-response camera tube as used here, such a large scan spot offers advantages. For example, during the recognition mode the sensor is more sensitive: in moving over the cathode, the image of the star remains for a correspondingly longer time within such a large scan spot. The accuracy with which the scan pattern holds the tracking-star image centrally is only slightly reduced by the use of the large scan spot.

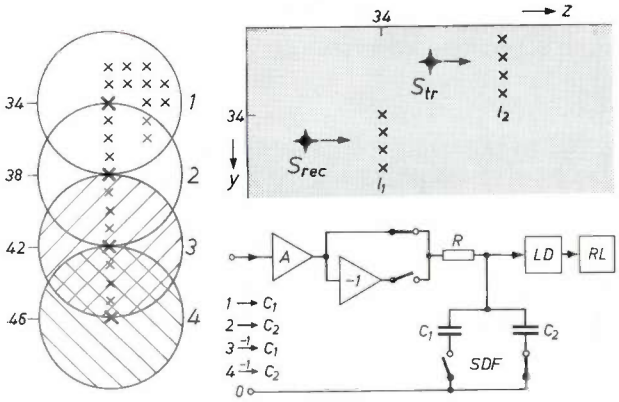


Fig. 3. The scanning of the photocathode (shaded) during the recognition mode. During this operation the images of the stars move with an angular velocity of 4 ± 2 minutes of arc per second (= 0.48 ± 0.24 mm/s) over the cathode. Before each measurement begins, the star sensor receives from the onboard computer the coordinate data for the scan lines l_1 and l_2 to which the images of the recognition star S_{rec} and the tracking star S_{tr} must be brought. Each scan line consists of 4 scan spots (*1*, . . . , *4*) with the coordinates (see diagram on left) $z = 34$, $y = 34, 38, 42, 46$. The crosses represent scan positions; for clarity, not all the possible positions are marked. The scanning of such a linetakes 1/64 s. The signal from a scan spot is fed to the sampled -data filter *SDF* via the video amplifier *A*, either direct or via an inverter (−1). The alternate opening and closing of the switches (whose control by the recognition logic *RL* is not shown) causes the capacitor C_1 to be charged by the signal current from scan spot *1* less that from scan spot *3*; similarly, the difference signal from scan spots *2* and *4* charges the capacitor C_2. The *RC* time constant of the filter is 0.3 s. If the star being searched for moves over the scan spot *3* of line l_1, then a (negative) voltage appears on C_1 which is large enough to trigger the level detector *LD*. The recognition logic then causes the second line to be scanned. If a star is also detected there within the next second, the two desired stars searched for have been found and recognized.

The recognition mode

For every astronomical observation the position data of a recognition star and of a tracking star are predetermined on Earth. The star sensor receives this data via the onboard computer. The satellite searches for the recognition star by rotating about the z-axis, which is pointed at the Sun, use being made of the data from the horizon sensor [4]. The telescope can then accurately scan a complete circle during each orbit around the Earth [1]. The images of the stars then move along 'horizontal' lines over the photocathode; thus only their z-coordinates change. To search for the star it is sufficient to continuously scan a short line on the cathode in the direction of the y-axis (*fig. 3*). The line consists of four scan positions — with three times three spaces in between — which together represent a sector of sky of 6 minutes of arc; this value is related to the accuracy with which the Sun pointing can be achieved [5]. Whenever a star of sufficient brightness passes the line, the video signal exceeds the threshold. The scan pattern then jumps to a second line to see whether the tracking star also passes over that. If a star is detected there within one second, the first star was indeed the recognition star desired, and the tracking mode commences. If no star is detected on the second line, the first star was not the recognition star; the scan pattern then jumps back to the first line and the search is continued. The series of operations during recognition is shown in the block diagram of *fig. 4*.

The tracking mode

In the tracking process a scanning pattern in the form of a cross is used; the star to be tracked is surrounded by four scan positions that are scanned in turn and touch in pairs (*fig. 5a*).

The control problem here is to keep the cross pattern continuously centred on the image of the tracking

star. In this way the onboard computer always has the coordinates of the tracking star available. This 'locking on' to the tracking star is done by the tracking-control system of the star sensor. The controlled quantity consists of the two coordinates (η,ζ) of the centre of the scan pattern; the image coordinates (y,z) of the tracking star form the set point of this control system. The error signal is derived from the video signal (fig. 5*b*). The correction takes place by changing the current in the deflection coils of the camera tube. A block diagram of the tracking-control system is shown in *fig. 6*. The system operates with a periodic sampling; the deviation between the coordinates (y,z) and (η,ζ) is determined every 1/64 second. Because this period is much shorter than the response time of the control system, the control is very nearly continuous [6].

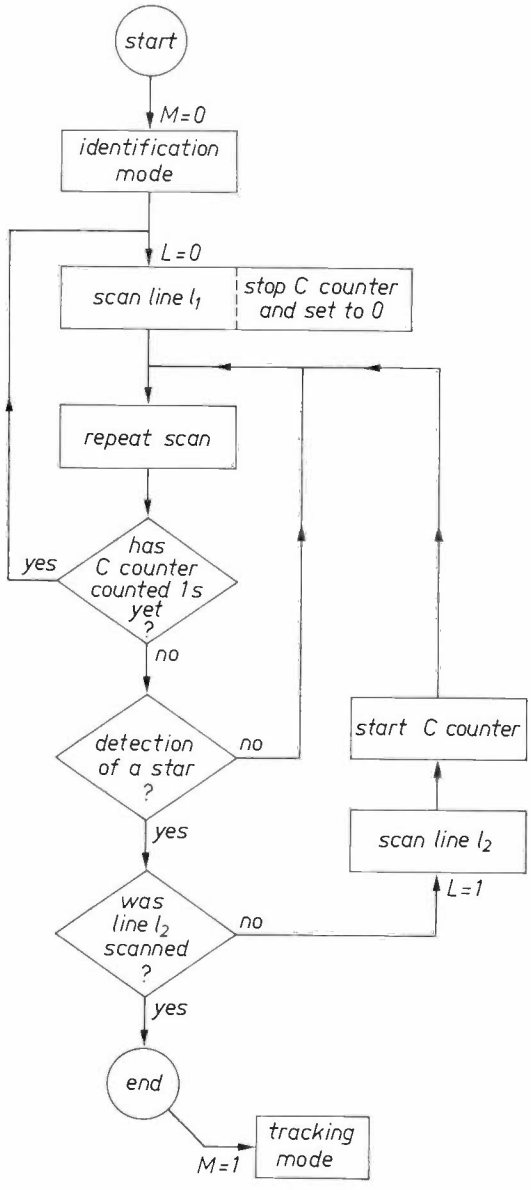


Fig. 4. Flow chart of the processes during the recognition mode. The status signals M and L are provided by the recognition logic (*RL* in fig. 2). The star sensor begins by working through the sequence of operations in the recognition mode. The coincidence counter C is started only when a star has been detected on the scanning line l_1 (see fig. 3). The conclusion that the two preselected stars (S_{tr} and S_{ree}, see fig. 3) have indeed been identified depends on the counter C; the second star must be detected on l_2 before C has counted one second. When M is changed the star sensor leaves the recognition mode and goes into the tracking mode.

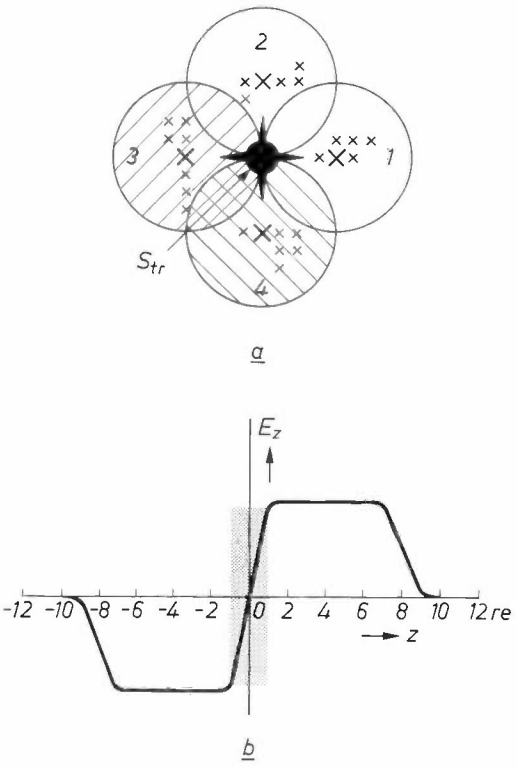


Fig. 5. *a*) During the tracking mode, the image of the tracking star S_{tr} on the photocathode is tracked by continually scanning four scan spots centred at the corners of a square in the sequence *1*, *2*, *3*, *4*. The diameter of the scan spot is equal to 8 times one resolution element; that of S_{tr} can be 2 resolution elements. Only a few of the possible scan positions on the photocathode are shown (crosses). The signal from the hatched scan spots is taken as negative, that from the other two as positive. When the tracking mode proceeds correctly, the pattern of the four scan spots will remain exactly symmetrical about the star image. *b*) The error signal E_z is obtained by subtracting the video signal from the scan spot *3* from the signal from *1* (similarly, the video signals from spots *2* and *4* yield the error signal E_y). In fig. 5*b* the signal E_z is plotted as a function of the difference between the set point (i.e. the coordinate z of the image of the tracking star) and the coordinate ζ of the centre of the scan pattern (fig. 5*a*). The value of this difference (the error) is expressed in terms of resolution elements (re). For a positive error, the signal from spot *1* is greater than that from *3*; for a negative error, just the reverse. In the centre E_z varies in proportion to $(z-\zeta)$; the proportionality factor increases with the brightness of the tracking star involved. If the image of the tracking star is exactly symmetrical about the centre of the scan pattern, the error signal is zero. The steep slope of the curve in the shaded area is a consequence of the fact that the star image in that area shifts over the edge of both scan spots; the diameter of the scan spot has no effect on the slope.

[4] See the article by P. van Dijk on the horizon sensor; this issue, p. 213.

[5] See the article by A. J. Smets on the sun sensors; this issue, p. 208.

[6] A discussion of the effect of the sampling period on control systems with periodic sampling is given in: J.-C. Gille, M. J. Pélegrin and P. Decaulne, Feedback control systems, McGraw-Hill, New York 1959, chapter 20.

The control is sufficiently rapid to respond to all expected movements of the satellite; the value of residual deviations will be an order of magnitude less than one resolution element, which corresponds to an attitude error of 21″. The compensating network or 'tracking filter' of fig. 6 consists of an integrator in series with a PI element (PI stands for 'proportional plus integrating'). Since the other elements give only constant gain, the control system is of the second order. In this way the scan pattern can follow the tracking star without any lag, even if the image moves at its maximum rate (6′ per second) over the photocathode. Without the PI element, a change of the set point at this rate would result in a constant lag of twice the resolution element [7].

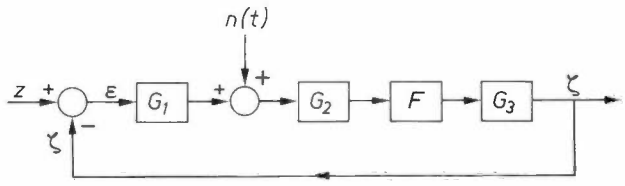


Fig. 6. Block diagram of the control system; ζ is the coordinate of the centre of the scan pattern (fig. 5*a*) to be controlled and z (= coordinate of the image of the tracking star on the photocathode) is its set point. For η, the other coordinate of the scan pattern, there is a similar control system with y as the set point. G_1 sensitivity of the sensor for position errors. G_2 amplification factor of the video amplifier. G_3 amplification factor of the deflection coils with their supplies. The external interference signal $n(t)$ is added to represent the noise effects in the camera tube, which cause fluctuation in the displacements of the scan pattern. F compensating network (tracking filter). This control system is of second order; in spite of the use of periodic sampling, the system works in effect continuously.

The open-circuit transfer function $H(j\omega)$ of the tracking-control system has the form $H(j\omega) = G_1G_2G_3\,(1 + 1/j\omega\tau_2)/j\omega\tau_1$, where G_1 is the sensitivity of the sensor for deviations in position (for the weakest tracking star — 8th magnitude — G_1 has the value of 0.6 nA per resolution element); G_2 is the amplification factor (in V/A) of the video amplifier and G_3 the amplification factor of the deflection system (in resolution elements/V). The ratio $(1 + 1/j\omega\tau_2)/j\omega\tau_1$ is the transfer function of the compensating network F. The time constants τ_1 and τ_2 are given values such that a star of the 8th magnitude can be tracked with an overshoot that is still acceptable (30%); τ_2 is 0.1 s and in the case mentioned $\tau_1/G_1G_2G_3$ is also equal to 0.1 s. For brighter stars the overshoot is smaller and the system also responds more rapidly. A few very bright stars cannot be used as tracking stars: the response would then be so rapid that the sampling rate would be too low, leading to instability. *Fig. 7* shows the behaviour of the coordinate $\xi(t)$ of the scan pattern during tracking. *Fig. 8* shows the circuit of the compensating network; the operation of the sampling by means of a number of switching transistors is also explained there.

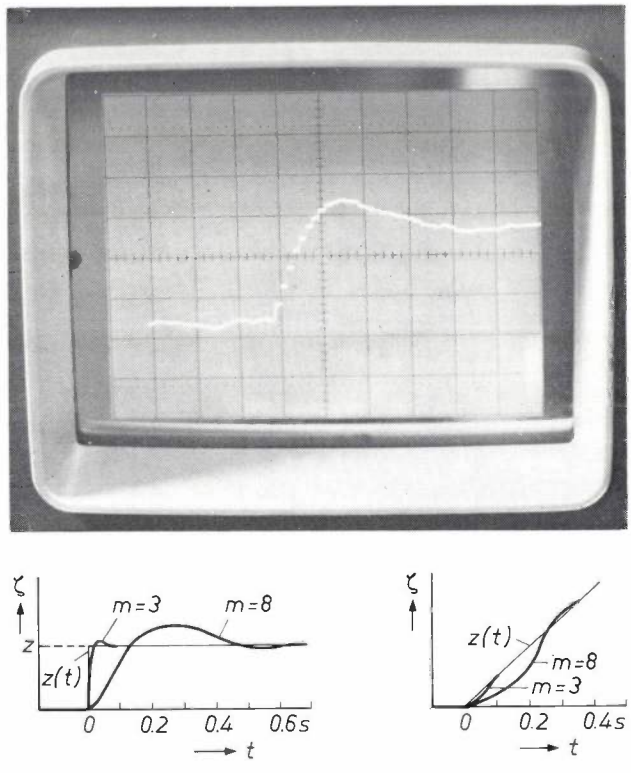


Fig. 7. Behaviour of the coordinate ζ of the scan pattern (fig. 3) during tracking as a function of time t. The coordinate $z(t)$ of the tracking star is the set point of ζ. During the tests on the tracking system, the set point was changed both stepwise and as a ramp function, for two brightnesses of the simulated tracking star. For a bright star, such as a star of the third magnitude ($m = +3$), the gain of the video amplifier is reduced. The photograph with the step response also shows, besides the overshoot that may be expected with faint stars, the sampling of the signal. It can be seen that, in spite of the signal sampling used, the response is effectively continuous. Because of the higher order of the control system, the response to the ramp function is such that the error $(z - \zeta)$ is rapidly reduced to zero.

The accuracy of tracking a star is of course subject to some limitation from the effects of various noise sources in the sensor, e.g. the statistical fluctuations in the emission of the photocathode in the camera tube. The bandwidth of the tracking-control system is purposely kept small (about 5 Hz); the time constants are thus large and the interfering effects of noise are then largely averaged out. The control is rather slow in response but it is fast enough to follow movements of the satellite. In the final section below on the prototype tests we enter into some further detail concerning the signal-to-noise ratio.

Prototype tests and results

In the course of the work on the satellite, three prototype versions of the star sensor were built. The first, the development model, was used in evaluating the design. Two almost identical versions were then made: the electrical prototype and the attitude-control prototype. The electrical prototype was used to investigate the operation of the sensor as a component of the complete satellite system, and the attitude-control prototype was

[7] See for example chapters 6 and 7 of the book of note [6].

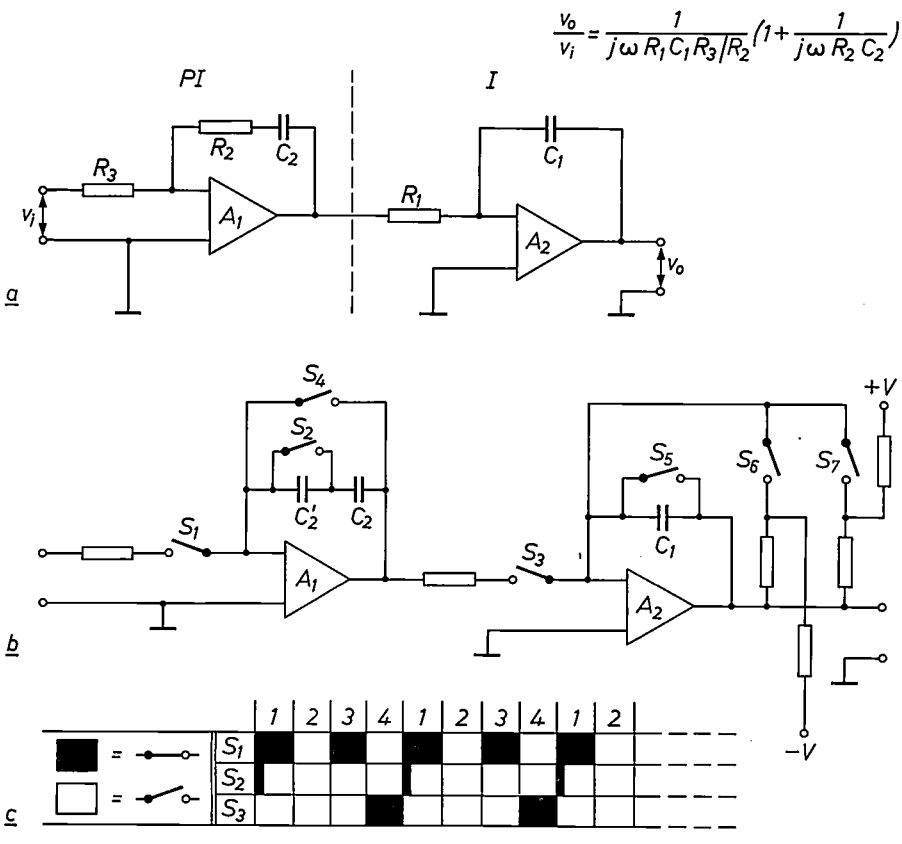


Fig. 8. *a*) Simplified block diagram of the compensating network (*F* in fig. 6) for one coordinate of the scan pattern for a continuous system. A_1, A_2 operational amplifiers with negligible input current and high gain. On the left of the dashed line is the PI element and on the right-hand side the integrator. The transfer function has the form $(1 + 1/j\omega\tau_2)/j\omega\tau_1$. *b*) As (a) but now operating with periodic sampling. The capacitor C_2' takes the place of the resistor R_2 in (*a*); the switches $S_1, \ldots, S_7$ are MOS transistors. S_4 and S_5 are closed until the tracking mode begins; all the capacitors are thus uncharged at that instant. S_1 isolates the control system of ξ from that of η. Because at the beginning of each scan cycle (*1*, *2*, *3*, *4*) the switch S_2 closes momentarily, C_2 is discharged. A voltage then appears across C_2' that is proportional to the difference signal of positions *1* and *3*. This signal is present when position *4* is being scanned: switch S_3 ensures that the I element only integrates then. The difference signal is integrated over many scanning periods due to the charging of C_2; the tracking error can thus be reduced exactly (and rapidly) to zero, even when the tracking star moves with a constant velocity over the photocathode (the ramp function, fig. 7). The switches S_6 and S_7 reset the I element (by means of voltages $-V$ and $+V$), as is necessary when the level detectors find position errors larger than one resolution element. Some of the switching operations are summarized diagrammatically in (*c*).

used to study star-sensor operation as part of the attitude-control system. The experience gained was then used in modifying the development model. Two flight models were built: one for the satellite and the other as a spare.

Investigation of the vibration effects expected during launching has shown that the star sensor will be able to withstand these without damage. The working of the sensor under various simulated conditions — vacuum, temperatures between −20 °C and 50 °C — has also been investigated in some detail. Some general data and characteristics of the sensor are given in *Table I*.

In the tests on the electrical and attitude-control prototypes, noise effects were investigated. The noise in the video signal is caused by the statistical character of the emission of both the cathode and the dynodes of the camera tube. A considerable contribution to the noise is made by two undesirable sources of light that give rise to emission: scattered sunlight and the weak background light of the innumerable stars of brightness less than eighth magnitude.

During the detection of stars in the recognition mode, the noise can give rise to various types of error. The video signal from a star that is being searched for

Table I. General data for the star sensor

Mass	2700 g
Power consumption	2.7 W
Permissible vibration levels	Accelerations up to 18*g* (for sinusoidal vibrations)
Temperature limitations	−20 °C to 50 °C
Field of view	1°30′×1°30′
Resolution element	21″
Star-image diameter	42″ (nominal)
Scan-spot diameter	2′48″
Scan rate	256 positions per second
Permissible apparent brightness	3rd to 8th magnitude[*]
Recognition mode	
Signal/noise ratio	15 (minimum)
Relative standard deviation of mean signal	7.5% (max)
Tracking mode	
Standard deviation of tracking noise	3″ (max)
Error due to changes in	
Temperature	< 0.2″/°C
Supply voltage	< 1″

[*] Stellar magnitudes are expressed on a negative logarithmic scale; a difference in magnitude of five on this scale represents a factor of 100 in brightness. A star of the 8th magnitude has a brightness (illuminance of 1.5×10^{-9} lumen/m^2 at the entrance pupil of the telescope; the light flux on the photocathode is then 2×10^{-11} lumen. Such a star is invisible to the naked eye; the faintest star that is just visible to the naked eye is of magnitude 5, 16× brighter. Bright stars are seldom available in the small field of view.

can be masked by the noise, bringing the signal below the threshold level of the level detector, so that the star is missed. The reverse is also possible; noise signals greater than the threshold can appear to be due to a star. In all these situations the shot noise of the photocathode has the largest effect.

In the worst case — a star of the eighth magnitude whose image passes over the scanning line in the shortest possible time (0.4 s) — the photocathode emits about 1600 electrons. This signal is superimposed on a fluctuating background, whose mean strength (d.c. component) corresponds to about 3200 electrons per scan position.

The background in this worst case thus gives a signal that is twice that of the star. In the output signal, however, the signal from the star is much greater than the background because the d.c. component is removed. This is done by reducing the signal by an amount equal to the signal originating in an adjacent scan position from which the star image cannot be seen. The *difference signal* then represents the new video signal (a subtraction circuit is already present in the system; a subtraction operation also takes place in the tracking mode (fig. 5) to produce the error signal).

The standard deviation of the amplitude of the video signal thus obtained is 90 electrons ($\sqrt{1600 + 2\times 3200}$; the shot noise has a Poisson distribution; the noise of the spurious radiation must be counted twice because of the subtraction). Taking into account the (small) noise contribution from the dynodes of the image-dissector camera tube and the efficiency (80%) of the sampled-data filter (fig. 3), the result found was that during the recognition mode the amplitude of the star signal has a relative standard deviation of not more than 7.5%. The worst-case value of the mean signal-to-background ratio is 15; the threshold of the level detector is set just half-way, making the probability of detection errors practically negligible.

During the tracking mode the scan pattern — again as a consequence of noise in the camera tube — will be subject to fluctuating displacements with respect to the star image (tracking noise). In the block diagram of fig. 6 the tracking noise is accounted for by adding the external noise signal $n(t)$. Because the power spectrum of the noise signal is flat up to frequencies well beyond the bandwidth of the system (white noise), the standard deviation of the displacements is given by:

$$\sigma_t = \frac{1}{G_1}\sqrt{\Phi_n B},$$

where Φ_n is the power per unit bandwidth of the noise signal $n(t)$ and B the bandwidth of the system.

Under worst-case conditions — tracking a star of the eighth magnitude against a background of the maximum brightness — the tracking noise has a standard deviation corresponding to only 15% of one resolution element.

A brief derivation will be given here of the formula for the standard deviation of the displacement fluctuations of the scan pattern [8] (for the z-coordinate only). The closed-loop transfer function giving the ratio between the noise signal $n(t)$ *as input signal* and the output signal $\zeta(t)$ is

$$Y_n(j\omega) = \frac{H(j\omega)}{G_1\{1 + H(j\omega)\}}.$$

G_1 appears in the denominator because the element G_1 in fig. 6 is located in the feedback branch, with respect to $n(t)$.

The output power per unit bandwidth is

$$\Phi_\zeta(\omega) = |Y_n(j\omega)|^2 \Phi_n,$$

where Φ_n, the input power per unit bandwidth, is independent of frequency.

The square of the standard deviation is

$$\sigma_t^2 \equiv \overline{\zeta^2(t)} = \frac{1}{2\pi}\int_0^\infty \Phi_\zeta(\omega)\mathrm{d}\omega.$$

Substituting the two previous expressions gives

$$\sigma_t^2 = \frac{1}{G_1}\Phi_n \frac{1}{2\pi}\int_0^\infty \left|\frac{H(j\omega)}{1 + H(j\omega)}\right|^2 \mathrm{d}\omega.$$

The integral in this expression is equal to 2π times the bandwidth B of the tracking-control system. The noise power Φ_n is that of the shot noise in the video signal which originates in the electron emission at the photocathode. The light from the tracking star and the background (mainly scattered sunlight) contribute to this noise.

Finally, it should be noted that the standard deviation is independent of the position of the image of the tracking star on the photocathode.

[8] See for example G. Quasius and F. McCanless, Star trackers and systems design, MacMillan, London 1966, chapter 10.

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 9

'MADGE', a microwave aircraft digital guidance equipment

I. General principles and angle-measuring units

R. N. Alcock, D. A. Lucas and R. P. Vincent

In a competitive series of NATO tests held in 1970 at Bedford, England, and in 1971 at Fréjus, France, the MADGE microwave aircraft-guidance equipment was declared the preferred system from a number of entries from Britain, France, W. Germany and the U.S.A. The tests at Bedford showed that MADGE could provide guidance equivalent to that now available at many airports, but over a wide range of angular coverage, and the tests at Fréjus showed that performance was not significantly affected by a change of sites. MADGE has an airborne transmitter and passive angle-measuring units on the ground; the ground equipment is readily portable and easy to set up quickly. The pilot can select the angle of approach and the system will give simultaneous guidance for a number of aircraft. The article below is the first of two [*] on the MADGE system, which is now in development for production.

Introduction

A guidance aid for aircraft landing defines a path in space terminating at a runway or landing pad. An aircraft that follows the path is guided to a safe position for landing. Visual guidance aids can be used over short ranges; arrays of lights on the ground are arranged in such a way that the pilot can tell from the pattern what corrections he has to make to get his aircraft on to the correct approach path. At longer ranges, and for conditions of poor visibility, guidance aids depending on radio signals have to be used. The signals can define a path typically 10-20 nautical miles (18-36 km) in length. With such an aid an aircraft can descend to some tens of metres above the ground without visual reference. Error information is presented to the pilot or autopilot as the deviation from the desired horizontal and vertical position.

The landing-guidance aid now in use at many airports throughout the world is the Instrument Landing System (ILS). This system guides the aircraft down to a height of typically 30 metres (100 ft) along a single path above the extended centre-line of the runway. The vertical angle between the ILS path and the centre-line of the runway is between 2.5° and 3°. The ground installation for ILS is relatively large and expensive, and the accuracy of the radio path in space may be dependent on the profile of the terrain or affected by large man-made structures.

The system to be described in this article [1], 'MADGE' (an acronym for Microwave Aircraft Digital Guidance Equipment), belongs to a new generation of microwave landing systems. These are smaller, less expensive to install, much less susceptible to the effects of ground profile and man-made obstacles, and can therefore be used at many more airfields than are currently served by ILS. Such microwave systems provide guidance along many approach paths at differing angles and simultaneously meet the requirements of a wide variety of aircraft types: fixed-wing, vertical and short take-off and helicopters.

The MADGE system has been designed to meet civil

R. N. Alcock, M.A., M.Sc., D. A. Lucas, B.Sc. (Eng.) and R. P. Vincent, M.A. are with Mullard Research Laboratories, Redhill, Surrey, England.

[*] The second article will appear in Vol. 35 of this journal. *(Ed.)*

[1] See also S. J. Robinson, Philips Telecomm. Rev. **32**, 155, 1974 (No. 3).

and military requirements for approach guidance and for navigation in the local area surrounding the airfield. The system consists of a number of modules and can easily be extended to meet new requirements in the future.

MADGE provides accurate guidance over wide angles of coverage: its angular accuracy (about 0.05°) is equivalent to that of a Category II standard ILS system. The pilot can select his desired approach angle; horizontal and vertical displacement errors and distance to the runway are displayed in the aircraft. The system is capable of providing simultaneous guidance for over 150 aircraft. The ground equipment, which is battery operated, is portable and can be set up quickly in a variety of temporary landing sites. Two men can set it up in 15 minutes.

Type of system

There are two basic methods of deriving angular information for guidance purposes, which differ in the location of the transmitter. In ILS, scanning-beam and doppler systems, a signal is generated by a *ground-based transmitter*. The signal is modulated by a code that indicates the direction in which it is transmitted. In the aircraft there is a special receiver that derives the positional information by decoding this signal. The information can be displayed on a cross-pointer meter that gives 'fly left', 'fly right', 'fly up' or 'fly down' instructions to the pilot.

MADGE is based on another kind of system, in which there is a *transmitter in the aircraft*. Each aircraft has its own microwave transmitter, which 'interrogates' a ground-based receiver system that measures the angular position of the aircraft. The information derived from this measurement is then instantaneously transmitted to the aircraft by a transponder forming part of a data link.

A system of this kind was chosen because it has three particularly attractive features, To begin with, the airborne equipment is essentially a data-link terminal, and little modification has to be made if at any time the angular coverage or accuracy of the system have to be changed.

Secondly, since modern systems are generally required to measure the slant distance from the landing site to the aircraft there must necessarily be some kind of two-way radio link from ground to air. This link can easily be made to carry the angular information as well, with no significant increase in complexity.

The other useful feature of such a system is that it is easy to extract the angular information on the ground, where it can be used in air-traffic surveillance and flight-path monitoring. Finally, deviations from a linear path can be calculated on the ground, so that the aircraft can be guided along curved flight paths if desired, without the need for expensive modifications in the airborne equipment.

The measurements of angle are made by *interferometers* [2]. This method is very accurate over a wide range of angular cover and is not greatly affected by reflection at the site. The signal processing in an interferometer can be readily altered to meet any change in requirements in the future.

General description of the system

A block schematic diagram of the system is shown in *fig. 1*. A microwave transmitter carried in the aircraft transmits a pulsed signal to the three angle-measuring receivers, which are interferometers that measure accurately the direction of arrival of the signal. Azimuth angles are measured by two identical interferometer units: one unit gives coverage for aircraft approaching the runway (*approach*), and the other gives coverage for aircraft carrying out a missed approach (*overshoot*). The elevation angle is measured in the approach sector only.

The angular information from the interferometer is fed to the *transponder*, a transmit/receive unit that sends the ground information back when it receives the interrogating signal from the aircraft. This angular information is presented to the transponder as a sequence of binary digits, called the 'angle word', and the transponder transmits this data back to the aircraft as an amplitude-modulated pulse train on a microwave carrier signal. A receiver in the aircraft decodes the angular information and compares it with azimuth and elevation angles that the pilot has selected on his control unit. The error signals that result from these comparisons are applied in analog form to conventional cross-pointer indicators as used in ILS. The slant distance between the aircraft and the landing site is derived digitally in the aircraft from the go-and-return time taken by the transmitted and return pulses and is presented in analog form to a meter in the aircraft. Differentiation of the analog output representing distance gives the velocity of the aircraft.

Both aircraft and ground transmissions carry coded address signals that identify the aircraft and the ground station. Guidance data received by the airborne equipment is only passed through to the indicating instruments when the air and ground addresses have been accepted in a validity check. This procedure enables the system to operate with a large number of aircraft and reduces susceptibility to interference. The signals sent out from the aircraft thus contain several distinct pulse groups. The complete pattern is called the 'interrogation word', and is repeated at a rate of 50 Hz. To avoid repetition interference between different aircraft the

word-repetition rate is randomly varied ('jittered') around the nominal value of 50 Hz. The first pulse in the interrogation word is the 'location' pulse, the one that is processed by the interferometers. This is followed by air and ground address codes, and parity bits for detecting errors. The pulses forming the interrogation word have a duration of 1 μs and define a bit rate of 1 MHz; the initial location pulse is 4 μs long. Similarly, operating frequency in the 5.00 to 5.25 GHz navigation band (C band).

The MADGE units are made up of modules: this means that the system can very easily be modified or rearranged to meet changing requirements. In particular the receiver modules of the angle-measuring units are easily rearranged. Changes in the ground equipment can be made without affecting the airborne equipment.

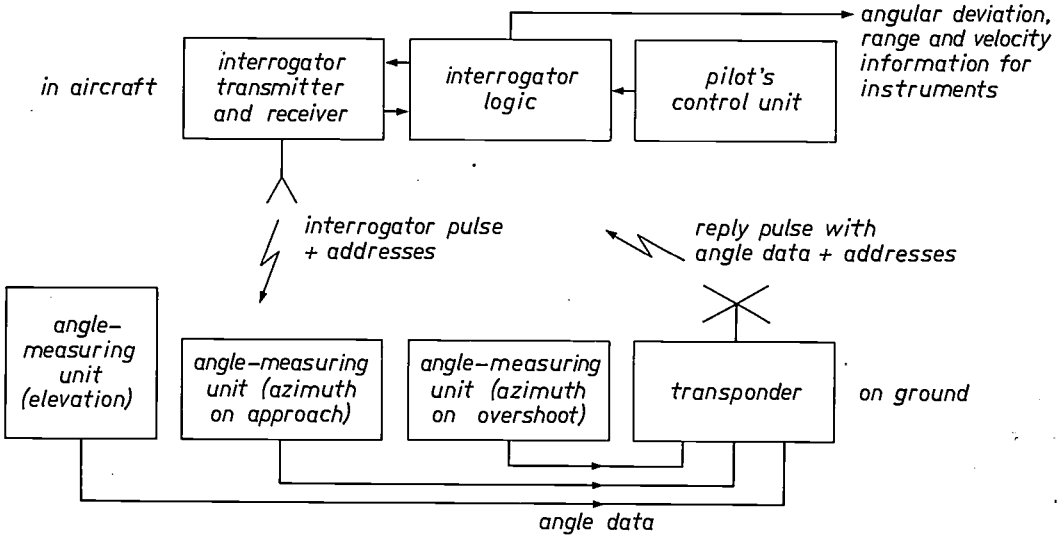


Fig. 1. Block schematic diagram of the MADGE system. A microwave transmitter in the aircraft transmits a pulsed signal to the three angle-measuring receivers, which are interferometers that measure accurately the direction of the signal. Azimuth angles are measured by two identical units: one for aircraft approaching the runway, the other for aircraft carrying out a missed approach (overshoot). Elevation angle is measured in the approach sector only.

The angle data from the interferometers is fed into the transponder, a transmit/receive unit that sends back the ground information in reply to the interrogating signal from the aircraft. This angle data is a series of binary digits, which the transponder transmits back to the aircraft as an amplitude-modulated pulse train on a microwave carrier signal. A receiver in the aircraft decodes the angle data and compares it with values selected by the pilot. The resulting error signals are applied in analog form to conventional ILS-type cross-pointer indicators. The slant distance from aircraft to landing site is derived digitally in the aircraft from the go-and-return times of the pulses and is presented in analog form to a meter. Differentiation of the analog output representing distance gives the velocity of the aircraft.

the reply word from the ground consists of a 'range' pulse, which is processed by the airborne distance-measuring equipment, followed by address codes, azimuth and elevation words, a fault-warning group and parity bits.

The choice of operating frequency is determined by two contradictory factors. As with any radio direction finder, increasing the size of the antenna array or the frequency increases the angular accuracy. A frequency in the microwave range will permit accurate angular measurements to be made with compact equipment. However, at higher operating frequencies more transmitter power is required to maintain the same angular coverage and to cope with atmospheric attenuation (mostly due to rain). The transmitter is therefore more difficult to design and more expensive. A good compromise can be obtained by locating the MADGE

Proposals for advanced landing systems that meet ICAO (International Civil Aviation Organization) and NATO requirements include a range of units giving straight-line guidance for all types of airfield and various weather limitations.

Since the data link can transmit the range to the ground station other useful facilities can be provided. Thus, the position and identity of all cooperating aircraft can be displayed at a ground station: this could simplify air-traffic control and allow flight paths to be monitored. Computations of deviation from correct path can also be made on the ground, enabling aircraft to be guided along nonlinear paths without expensive modifications to the airborne installation.

[2] R. N. Alcock, A digital direction finder, Philips tech. Rev. 28, 226-230, 1967.

Table I summarizes the coverage and accuracy of the system. The design and operation of the angle-measuring interferometers will now be described. A description of the data link and distance-measuring equipment will be presented in a subsequent article.

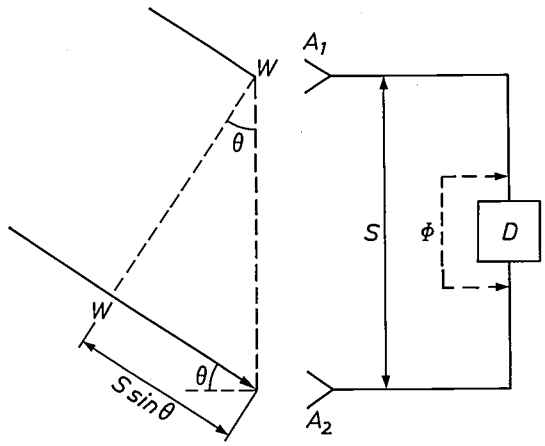


Fig. 2. Measurement of angle with an interferometer. The plane wavefront WW meets the line joining the antennas A_1 and A_2 at an angle θ. The two antennas are separated by a distance S and connected to a phase discriminator D. The phase difference Φ depends on the angle θ.

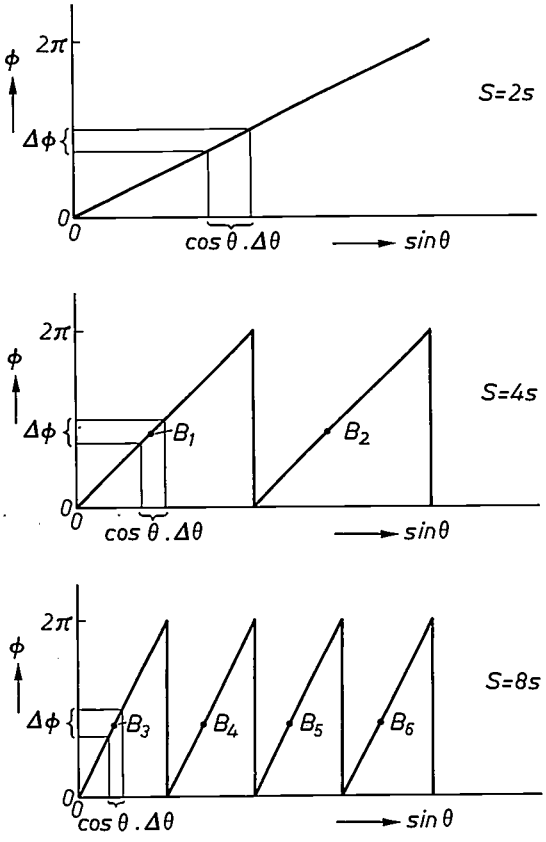


Fig. 3. Variation of measured phase difference ϕ with angle for antenna spacings $S = 2s$, $4s$ and $8s$ as in the azimuth unit. On increasing the spacing the accuracy increases since $\cos\theta.\Delta\theta$ is smaller for the same value of $\Delta\phi$. However, when the path difference is greater than the wavelength, ambiguity arises as at B_1 and B_2 or B_3 to B_6). These ambiguities can be resolved by reference to phase measurements from less widely spaced antenna pairs.

Table I. Coverage and accuracy of the MADGE system.

A. Coverage

Measured quantity	Horizontal	Vertical	Range (km)
Distance (data link)	360°	1° to 40°	27
Azimuth angle θ	90° (130° at close range)	1° to 40°	27
Elevation angle β	90°	1.5 to 25°	27

B. Accuracy

Measured quantity	Variations over periods > 15 min (long-term stability)	Variations over periods < 15 min (dynamic noise)
Distance		
at 2 km	6 m	3 m
at 27 km	35 m	47 m
Azimuth angle θ	0.08°	0.03°
Elevation angle β	0.08°	0.03°

Measurement of angle by multiple interferometers

The angle-measuring units are built up from a number of simple interferometers. *Fig. 2* shows a simple interferometer consisting of two antennas separated by a distance S. The plane wavefront WW is assumed to originate from a distant transmitter at an angle θ relative to the normal to the line joining the two antennas, and the phase difference is measured in a discriminator D as shown. The phase difference Φ of signals *entering* the discriminator is given by:

$$\Phi = \frac{2\pi S}{\lambda} \sin\theta, \qquad (1)$$

where λ is the signal wavelength. Thus, if the phase difference Φ is measured and the antenna spacing and signal wavelength are known, the angle can be determined.

The rate of change of phase with angle is given by:

$$\frac{d\Phi}{d\theta} = \frac{2\pi S}{\lambda} \cos\theta. \qquad (2)$$

This shows that for a particular phase-measuring accuracy the accuracy of the measurement of angle increases with spacing. *Fig. 3* illustrates the increase in accuracy as the spacing is increased. It also shows that phase ambiguity occurs as the spacing is increased. This arises because the discriminator only measures phase difference between 0 and 2π, taking no account of integer multiplies of 2π, which are implicit in equation (1). The closely spaced interferometer is unambiguous but relatively inaccurate, whereas the widely spaced interferometer is ambiguous but accurate.

In a multiple interferometer the inherent ambiguity of the widely spaced pair can be resolved by reference to the phase measurements made by the less widely spaced pairs. In this way a high angle-measuring accuracy may be achieved unambiguously over a wide angle of coverage.

Choice of spacings

Various interferometer arrangements can be used to resolve the ambiguities. We shall now consider some general arguments relating to ambiguity resolution.

Fig. 4 shows two interferometers k and $k+1$ with spacings S_k and $S_{k+1} = nS_k$. These form part of a set of interferometers with progressively increasing spacing. The unambiguous signal phase difference Φ_k between signals entering interferometer k is composed of the phase difference ϕ_k measured by the discriminator (ϕ_k lies between 0 and 2π) and an integer multiple of 2π, so that:

$$\Phi_k = \phi_k + 2\pi I_k \qquad (3)$$

and

$$\Phi_{k+1} = \phi_{k+1} + 2\pi I_{k+1}, \qquad (4)$$

where I_k and I_{k+1} are integers.

We consider a situation in which the phase difference Φ_k is known and the phase difference Φ_{k+1} has to be determined. It is assumed that I_k is known, ϕ_k and ϕ_{k+1} are measured by the discriminators and I_{k+1} is unknown. From (1):

$$\Phi_k = \frac{2\pi S}{\lambda}\sin\theta \quad \text{and} \quad \Phi_{k+1} = \frac{2\pi nS}{\lambda}\sin\theta,$$

so that

$$I_{k+1} = \frac{1}{2\pi}\left\{n(\phi_k + 2\pi I_k) - \phi_{k+1}\right\}. \qquad (5)$$

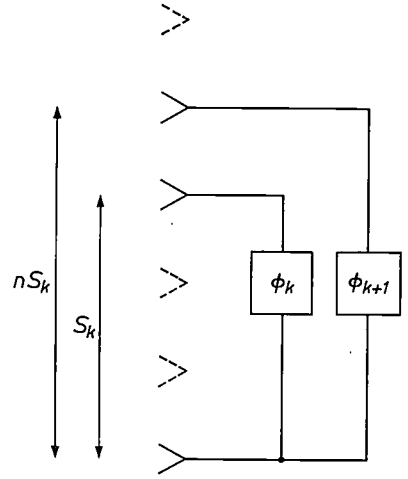


Fig. 4. Two successive pairs k and $k+1$ from a set of interferometers, with spacings S_k and $S_{k+1} = nS_k$. The corresponding measured phase differences are ϕ_k and ϕ_{k+1}.

This equation forms the basis for an iterative solution; if I_k is known it enables I_{k+1} to be determined. Starting with a closely spaced unambiguous interferometer for which $I_1 = 0$, I_2 can be determined for the next interferometer and so on up to I_{k+1}. The angle θ can then be derived by way of eqs. (4) and (1).

In practice, of course, there will be some uncertainty in the measurement of the phases. Uncertainty can arise from multipath propagation effects, residual instrumental inaccuracies and the quantization error that arises when phase differences are digitized. It is important to know how large these uncertainties may become before ambiguity resolution becomes impossible. Let us assume that each phase measurement includes a worst-case error e radians, which is the same for the pairs k and $k+1$, and is additive. Now the relation (5) will not give the integer I_{k+1}, but instead a quantity H, where

$$H = \frac{1}{2\pi}\left\{n(\phi_k + e + 2\pi I_k) - \phi_{k+1} + e\right\}. \qquad (6)$$

I_{k+1} is taken as the nearest integer to H, and the correct value will only be obtained if

$$|H - I_{k+1}| < \tfrac{1}{2}.$$

From eqs. (5) and (6) this may be written:

$$\frac{1}{2\pi}(n+1)\,|e| < \tfrac{1}{2},$$

or

$$|e| < \frac{\pi}{n+1}. \qquad (7)$$

Thus we have:

n	1.5	2	3	4
$\lvert e\rvert_{\max}$	$\pi/2$	$\pi/2.5$	$\pi/3$	$\pi/5$.

In practical interferometer systems the phase error e can approach a value of $\pi/3$, and the value of n should therefore not exceed 2. Logical processing techniques are also simplest when n is an integer.

Methods of processing using values of n greater than unity are applicable to a particular class of interferometers in which the spacings increase in geometric progression; the coverage is determined by the closest-spaced pair and the permitted phase uncertainty is independent of the bearing of the incident signal.

Another useful form of array is one in which spacings increase by equal increments to form a linear array. In this case it is found to be convenient to process phases according to equation (5) with the substitution $n = 1$. The permitted phase uncertainty may then be shown to be a function of the angle θ, reaching a maximum value of $\pi/2$ in a direction at right angles to the line of the array (where $\theta = 0$).

The design of the azimuth array is based on interferometers with spacings increasing in geometric pro-

gression with ratio 2. The spacings in the elevation array increase by equal increments forming a linear array. The phase-processing methods, which will be described for both arrays, are both iterative but differ significantly in the implementation.

In our description of the measurement of angle by multiple interferometers we have as yet only said that the phase differences are measured in discriminators. In fact these discriminators form part of the phase-measuring receivers, which will be discussed in the next section.

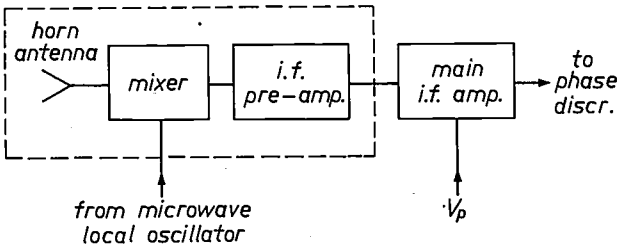


Fig. 5. Schematic diagram of one of the microwave receivers of an interferometer system. The antenna is included with the mixer and i.f. preamplifier to form a single sub-assembly. There is a single local oscillator for the complete array. V_p variable voltage applied for phase adjustment.

The interferometers

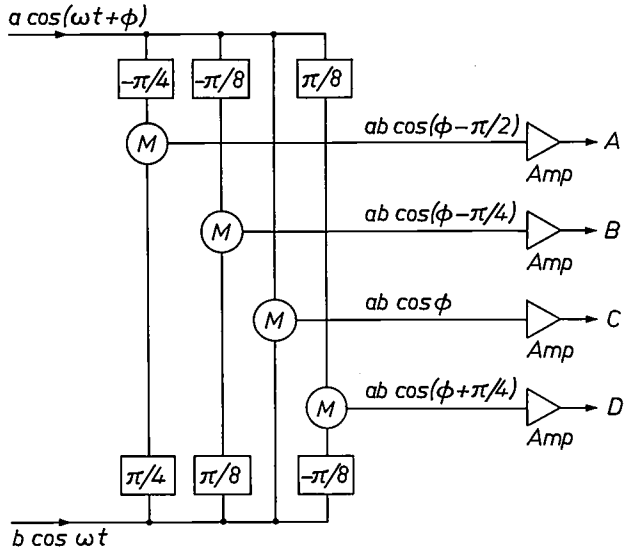
The interferometers are built up from numbers of appropriately spaced horn antennas, superheterodyne microwave receivers and phase discriminators, which form the building blocks of the system. The microwave phase difference is preserved at the *intermediate frequency* (i.f.) in the receiver, and the phase discriminators operate at i.f. We shall now look more closely at the microwave receiver and the i.f. phase discriminator.

The microwave receiver

A schematic diagram of one of the microwave receivers with its antenna is shown in *fig. 5*. Signals received in the horn antennas are mixed and amplified in a superheterodyne receiver with an intermediate frequency of 12 MHz. There is a single crystal-controlled local oscillator for each complete interferometer array. The horn antenna, microstrip balanced mixer and i.f. preamplifier form a sub-assembly. The preamplifier determines the i.f. bandwidth, which is 3 MHz, and the output signal from the preamplifier is applied to a relatively broadband main amplifier. This is a limiting amplifier and includes an electronic phase-shifter with a range of 45° for phase adjustment when the unit is set up. The mixer and amplifiers operate over a wide dynamic range at the r.f. signal input (about 70 dB). The signal phase is not significantly affected by the change in input level or by i.f. drift of up to 1 MHz. The output signal from the i.f. amplifier is applied to one of the phase-discriminator inputs.

The i.f. phase discriminator

The i.f. phase discriminator, shown in the schematic diagram of *fig. 6a*, is an arrangement of four broadband video multipliers M. Each multiplier provides output voltages proportional to the product of the amplitudes of the two input signals, and also to the cosine of the phase difference. Thus with input signals $a\cos(\omega t+\phi)$ and $b\cos\omega t$ the output voltage is proportional to $ab\cos\phi$). The input signals for one of the multipliers are taken directly from the i.f. outputs



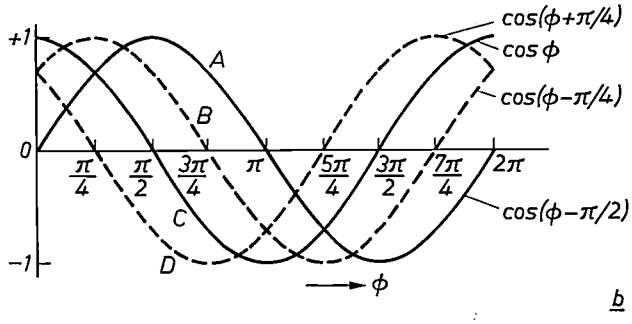


Fig. 6. *a*) Schematic diagram of one of the i.f. phase discriminators of an interferometer system. M broadband multiplier. The rectangular boxes represent phase shifters. The input signals $a\cos(\omega t+\phi)$ and $b\cos\omega t$ are taken from the two i.f. outputs of an interferometer pair, and are applied directly or through various phase shifters to the multipliers. Each multiplier provides output voltages proportional to the product of the input amplitudes, and also to the cosine of the phase difference. The phase shifters are given values such that the output voltages produced are proportional to $ab\cos(\phi-\pi/2)$, $ab\cos(\phi-\pi/4)$, $ab\cos\phi$ and $ab\cos(\phi+\pi/4)$. These voltages (see fig. 6*b*) are very suitable for digization of the measured phase angle. This takes place in the four comparator amplifiers *Amp*, which sense only whether the input voltage is positive or negative. A, B, C, D digitized outputs. *b*) The voltage waveforms applied to the input of the comparator amplifiers *Amp*, labelled with the appropriate output (see Table II). Thus for the range $0<\phi<\pi/4$ A, B, C and D are all positive, giving the code 0000, whereas for $\pi/4<\phi<\pi/2$ D has become positive, giving the code 0001.

of the two channels of an interferometer pair. The same input signals are applied to the other multipliers, but through pairs of fixed phase shifters that give total phase shifts of $-\pi/2$, $-\pi/4$ and $\pi/4$. The four output voltages are proportional to $ab\cos(\phi-\pi/2)$, $ab\cos(\phi-\pi/4)$, $ab\cos\phi$ and $ab\cos(\phi+\pi/4)$. These quantities are shown in fig. 6*b* as a function of the phase difference ϕ of the interferometer pair; the phase shifts have the particular values chosen because the output voltages are then suitable for *digitizing* the measured phase angle. This takes place in the four comparator amplifiers shown as *Amp*, which sense only whether the input voltage is positive or negative. For a positive input the output voltage is at logic level '0', and for a negative input the output is at logic level '1'. Thus for the range $0<\phi<\pi/4$, fig. 6*b* shows that the outputs *A*, *B*, *C* and *D* are all positive, which can be expressed by the code 0000, whereas in the range $\pi/4<\phi<\pi/2$ the output *D* has become negative, giving the code 0001. Consequently the value of ϕ can be expressed in steps of $\pi/4$ in the code shown in *Table II* (this kind of code is known as a Johnson code).

Table II. Digital outputs from the discriminator for various ranges of the input phase difference ϕ. (The successive outputs *ABCD* form a Johnson code.)

ϕ	Discriminator outputs			
	A	*B*	*C*	*D*
0 - $\pi/4$	0	0	0	0
$\pi/4$- $\pi/2$	0	0	0	1
$\pi/2$-$3\pi/4$	0	0	1	1
$3\pi/4$- π	0	1	1	1
π -$5\pi/4$	1	1	1	1
$5\pi/4$-$3\pi/2$	1	1	1	0
$3\pi/2$-$7\pi/4$	1	1	0	0
$7\pi/4$-2π	1	0	0	0

A feature of this arrangement is that the crossover points of fig. 6*b*, corresponding to the digital steps, are independent of the relative levels *a* and *b* of the input signals. An extension of this technique is also used to digitize phase into increments of $\pi/8$. This uses eight multipliers and appropriate phase shifters.

Before going on to describe the operation of the units that measure azimuth and elevation angles, we should mention that there is *automatic self-monitoring* of the receiver and discriminator operation. Built-in test equipment generates a sequence of pulsed i.f. signals, which are injected into the i.f. amplifiers. The self-monitoring system detects failures in the receivers and discriminators and provides a warning both on the ground and in the aircraft.

The azimuth unit

The basic design of the azimuth unit is illustrated by the block diagram of *fig. 7*. It has a horizontal array of six simple interferometers, which share the common reference channel *C*. The spacings of five of these interferometers increase in a geometric progression of ratio 2.

In principle, ambiguities are resolved in the way illustrated in fig. 3. For example the two ambiguous azimuth angles corresponding to B_3 and B_4 are resolved by reference to the phase difference on the 4*s* interferometer; ambiguous bearings B_1 and B_2 on this interferometer are resolved by reference to the phase difference on the 2*s* interferometer. The geometric progression of antenna spacing has the advantage that large baselines, and hence high accuracy, can be built up with a relatively low number of channels.

It can be seen from fig. 7 that there is no pair of antennas with the spacing *s* for resolving ambiguities from the 2*s* pair. Equation (1) shows that such an *s* pair should have a maximum spacing of $\lambda/2$ to give unambiguous angular coverage of 180°, but with the preferred design of antenna this is physically impossible for a wavelength of 6 cm. The difficulty has been avoided by including a 3*s* pair and resolving ambiguities from the 2*s* pair. The value of *s* is 2.75 cm.

Horns are used as the antenna elements since in practice their phase characteristics remain closely similar as the angle or frequency is varied (good phase-tracking). The azimuth elements have a horizontal beamwidth of 70° and a vertical beamwidth of 26° (these are half-power beamwidths). They are tilted upwards at an angle of 15° to reduce the effect of ground reflection.

The arrangement of the receivers *R* and the discriminators *D* can be seen in fig. 7. The receiver in the common channel *C* drives all the discriminators via a seven-way splitter. The discriminator connected to the widest-spaced pair 32*s* has eight multipliers and digitizes the phase difference in increments of 22½°. All the other discriminators have four multipliers, and digitize the phase difference in increments of 45°. A 'threshold unit' is also connected to the common channel; this unit senses the strength of the received signal, and if it is above a certain minimum level the discriminator output signals are sampled and held in the logic unit for the duration of the logical processing (about 10 μs). The phase modulator *PM* provides a 6 Hz triangular modulation of the phase difference from the widest-spaced pair (32*s*). This phase modulation gives an interpolation within the digital interval of 22½°; we shall return to this later.

A separate sector-definition unit, with three antennas at approximately 120°, each with its own receiver, is

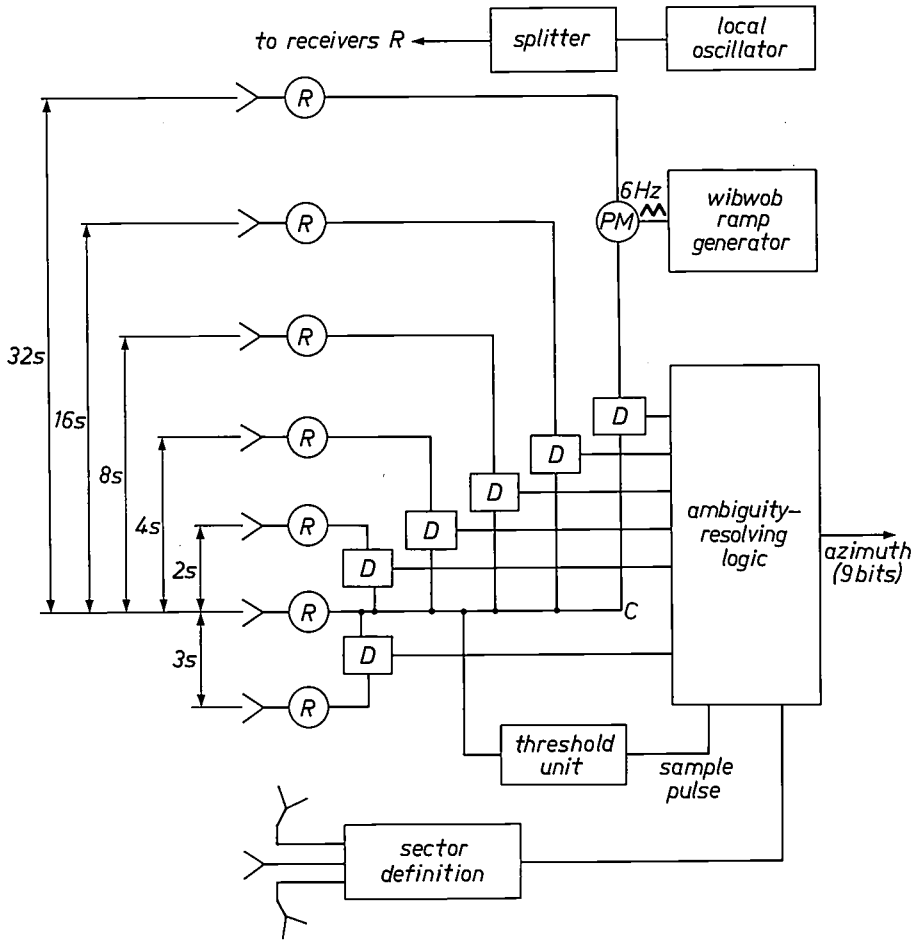


Fig. 7. Schematic diagram of the azimuth unit. There are six simple interferometers, sharing the common reference channel *C*. The spacings of five of the interferometers increase in a geometric progression of ratio 2. The unit of antenna spacing *s* is equal to 2.75 cm. *R* receiver. *D* discriminator. *PM* phase modulator. The lower group of three antennas with receivers is a separate sector-definition unit to show independently whether signals lie within a $\pm$ 65° sector on the approach side.

provided to give an independent indication of whether signals lie within a 130° sector on the approach side. This unit functions by comparing signal amplitudes and prevents the system from operating and showing unreliable angles when the aircraft is outside this sector. It also enables a distinction to be made between signals from the 'approach' and 'overshoot' sectors.

Processing the azimuth-angle data

The ambiguity-resolution process, outlined earlier, is carried out by logic circuits that essentially act as a 'look-up table' (i.e. output codes are selected for each specific input-code combination). The digital phase states of the discriminator are examined in appropriate groups and an unambiguous pure-binary output code is generated. The logic circuits (the 'processing logic'), are divided into discrete functional modules, shown in the block diagram of *fig. 8*.

The output signals from the discriminator *D* of the widest-spaced interferometer (fig. 7) are available as a parallel eight-digit Johnson code. These signals are converted directly into a pure-binary code forming the four least-significant bits B_6, B_7, B_8 and B_9 of an azimuth-angle word. The value of the B_9 bit therefore represents $22\frac{1}{2}$° of the phase difference ϕ_{32s} from the widest interferometer. The next interferometer has spacing $16s$ and the measured phase difference ϕ_{16s} is digitized to 45°; the Johnson-code output from the $16s$ discriminator is used directly in the ambiguity-resolution process. This is illustrated in *Table III*, which shows the combination of the Johnson code with the pure-binary outputs B_7 and B_6 to generate B_5, the next binary digit in the azimuth-angle word. An extra $\pi/8$ phase shift is provided in the discriminator so that the output quantity corresponds to $\phi_{16s} + \pi/8$; this ensures that the cross-over points of B_7 and the *ABCD* code are symmetrically related. What Table III shows is in effect a 'truth table' that for given values of B_6 and B_7 shows which of the digits *A*, *B*, *C* or *D* should be examined to provide the correct digit for B_5 in the azimuth-angle word. In formal logic notation, the operation of the truth table is expressed by the relation

$$B_5 = \bar{B}_6.B_7.A + \bar{B}_6.\bar{B}_7.B + B_6.B_7.\bar{C} + B_6.\bar{B}_7.\bar{D}. \quad (8)$$

In the table, the values to be examined are shown in the 'boxes'. Thus, if $B_6B_7 = 01$ and $ABCD = 1110$, then $B_5 = 1$.

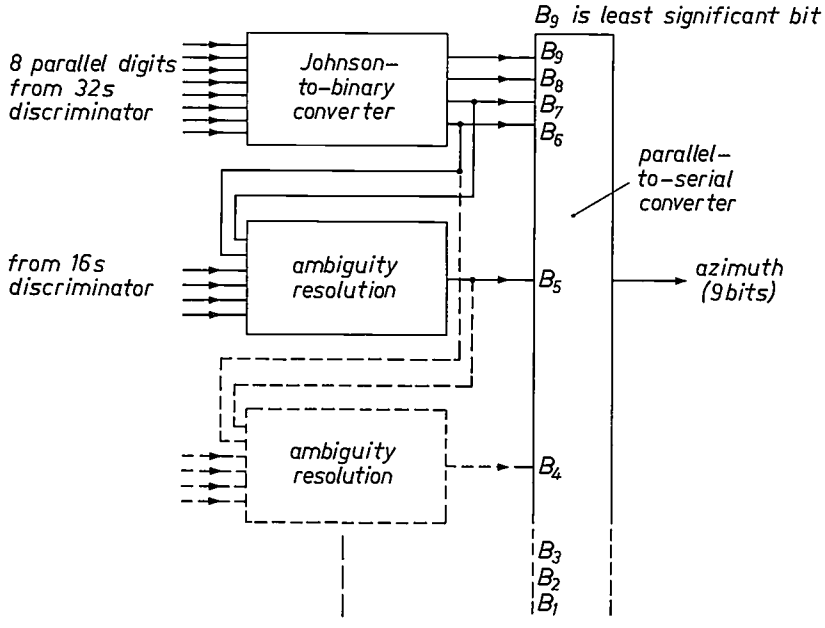


Fig. 8. Block diagram of the processing logic for resolving ambiguous angle measurements from the azimuth array. The process is carried out by logic circuits that essentially act as a 'look-up' table. Digital phase states of the discriminator are examined in appropriate groupings and an unambiguous pure-binary output is generated, as illustrated in Table III.

These operations are performed in one of the ambiguity-resolution units. Similar operations are performed for the other pairs to give the digits B_4 to B_1. The truth table is designed in such a way that the values of A, B, C and D that are examined fall in the most certain range of ϕ_{16s}, i.e. as far as possible from a transition. In fact, the table shows that a phase error of as much as $\pm 67\frac{1}{2}°$ is permitted on the 16s interferometer before ambiguity resolution breaks down. It can be shown that this is equivalent to a maximum permitted error of $\pm 45°$ in the measurement of ϕ_{16s} and ϕ_{32s}. The end product of the logic circuits is a group of nine parallel binary digits, which are converted into a time sequence (serial form) suitable for transmission over the data link to the aircraft.

Table III. Truth table corresponding to the logic equation $\bar{B}_5 = B_6 \,.\, B_7 \,.\, A + \bar{B}_6 \,.\, \bar{B}_7 \,.\, B + B_6 \,.\, B_7 \,.\, \bar{C} + B_6 \,.\, \bar{B}_7 \,.\, \bar{D}$

$\phi_{16s} + \pi/8$	B_6	B_7	16s Johnson code A	B	C	D	B_5
$\pi/8$- $\pi/4$	0	0	0	[0]	0	0	0
$\pi/4$- $3\pi/8$	0	0	0	[0]	0	1	0
$3\pi/8$- $\pi/2$	0	1	[0]	0	0	1	0
$\pi/2$- $5\pi/8$	0	1	[0]	0	1	1	0
$5\pi/8$- $3\pi/4$	1	0	0	0	1	[1]	0
$3\pi/4$- $7\pi/8$	1	0	0	1	1	[1]	0
$7\pi/8$- π	1	1	0	1	[1]	1	0
π - $9\pi/8$	1	1	1	1	[1]	1	0
$9\pi/8$- $5\pi/4$	0	0	1	[1]	1	1	1
$5\pi/4$-$11\pi/8$	0	0	1	[1]	1	0	1
$11\pi/8$- $3\pi/2$	0	1	[1]	1	1	0	1
$3\pi/2$-$13\pi/8$	0	1	[1]	1	0	0	1
$13\pi/8$- $7\pi/4$	1	0	1	1	0	[0]	1
$7\pi/4$-$15\pi/8$	1	0	1	0	0	[0]	1
$15\pi/8$- 2π	1	1	1	0	[0]	0	1
0 - $\pi/8$	1	1	0	0	[0]	0	1

The digital method we have outlined gives a phase resolution of $22\frac{1}{2}°$. However, as we mentioned earlier, it is possible to interpolate within this digital interval by means of a low-frequency phase modulation. This technique, known as *wibwob*, interpolates within the digital interval over several cycles of the modulation. The measured phase difference ϕ_{32s} is modulated at 6 Hz in the phase modulator *PM* (see fig. 7) by a triangular voltage waveform from the ramp generator. *Fig. 9a* shows that the variation with time of the modulated phase difference ϕ_t, which is given a peak-to-peak value of $22\frac{1}{2}°$, in relation to two digitization boundaries separated by $22\frac{1}{2}°$. The value of the unmodulated phase difference is of course ϕ_{32s}, and the modulated phase difference ϕ_t in general swings into the range corresponding to the next digital unit of $22\frac{1}{2}°$. The digital angular data now available for transmission to the aircraft therefore has a least-significant digit whose value changes when the ϕ_t curve runs over the boundary B. When the digital angular data is received at the aircraft it is converted into analog form to suit the aircraft instrumentation. Fig. 9*b* shows the unsmoothed analog waveform (solid line) that would result from this process; the pulses result from the periodic changes in value of the least-significant digit. In practice this waveform is smoothed by a lowpass filter inherent in the aircraft instrumentation (time constant 0.5 s), giving the smoothed signal (dashed line). This smoothed signal interpolates between the digital-boundary levels, giving an effective phase resolution of about 2°.

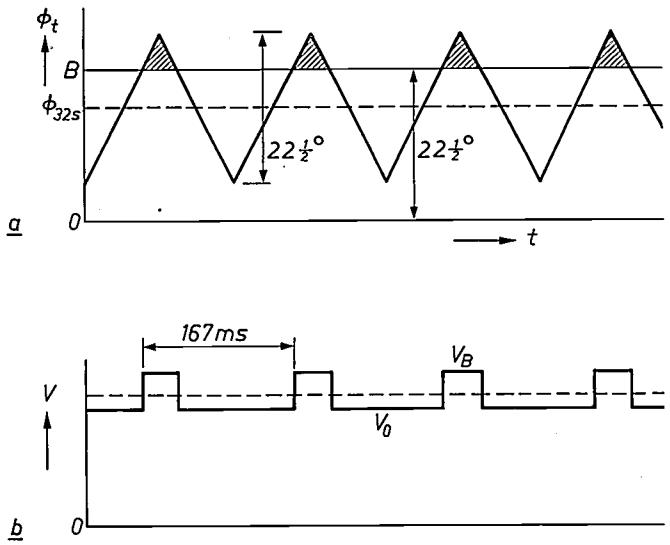


Fig. 9. Diagrams to illustrate the 'wibwob' technique, which increases the phase resolution. *a*) The measured phase difference from the widest-spaced pair (ϕ_{32s} for the system of fig. 7) is modulated at 6 Hz to give the triangular waveform ϕ_t shown. The peak-to-peak value of ϕ_t is 22½°, which is also the separation between the digitization boundaries *B* and 0. The digital angular data is transmitted to the aircraft, and converted to analog form. *b*) The solid curve shows the unsmoothed analog waveform available in the aircraft, and the dashed curve shows the smoothed waveform obtained after a lowpass filter with a time constant of 0.5 s (inherent in the aircraft instrumentation). The smoothed waveform interpolates between the digital-boundary levels, giving an effective phase resolution of about 2°.

The elevation unit

Ground reflections

Before going on to discuss the array for measuring the angle of elevation β, we should say something about the nature of the wavefront perturbations caused by ground reflections. These do not introduce any significant errors in the measurement of azimuth angles, because the antennas of the azimuth array are all at the same height and the errors cancel in the phase-difference measurements. However, the antennas of the elevation array are mounted in a *vertical* line, and the errors do not cancel. *Fig. 10* shows the patterns of amplitude and phase perturbation that result from the interference at a point *P* at a height *h* of a direct wave W_1 and a wave W_2 that has been reflected from the ground. At angles of incidence below about 3° the reflection coefficient of most surfaces approaches unity and the amplitude interference pattern contains deep nulls, associated with sharp phase discontinuities that approach 180°.

Design principles

Because of the phase and amplitude variations that occur in the vertical plane of the elevation array, the design approach is rather different from that used for the azimuth array. An array with 2 : 1 spacing ratios cannot be used, for two reasons. In the first place, the phase errors at the array (see fig. 10) would be too large to permit correct ambiguity resolution (see eq. (7) and p. 231). Secondly, for an array with 2 : 1 spacing ratios the angular accuracy is determined entirely by the widest-spaced pair. (The others only resolve the ambiguity.) With the kind of phase perturbation shown in fig. 10 any such single pair could be subject to a very large phase error if the perturbation was positive at one antenna and negative at the other. However, if more points on the wavefront are sampled along the baseline, the direction of the direct wave may be determined by averaging the wavefront perturbation, provided that the amplitude of the reflected wave is less than that of the direct wave. The difference between the direct and reflected amplitudes received by the antennas increases with the elevation angle β, partly because the ground reflection coefficient decreases with increasing angle of

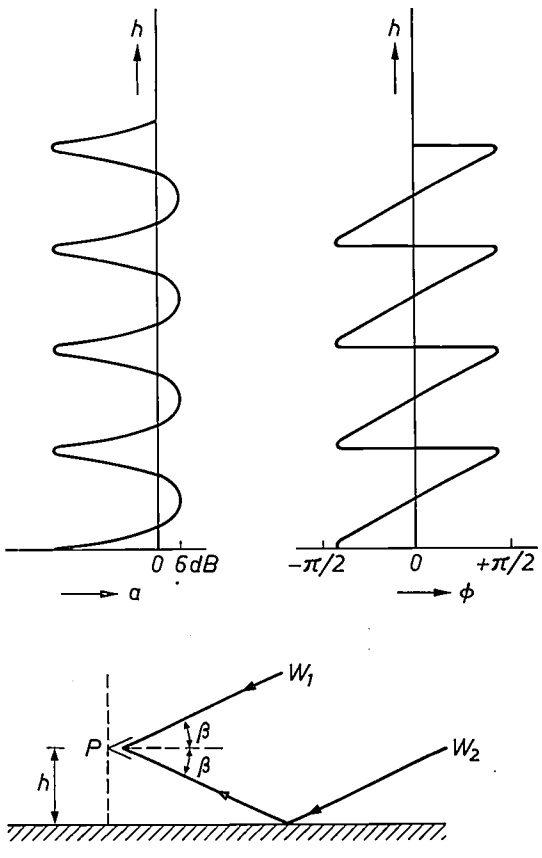


Fig. 10. When signals from an aircraft at an elevation angle β arrive at a vertical antenna array there is intereference between the direct wave W_1 and the wave W_2 reflected from the ground. The left-hand diagram shows a typical variation of amplitude *a* with height *h* of the point *P* above the ground. The right-hand diagram shows a typical variation of the measured phase ϕ with *h*. The phase variations encountered are too large for correct operation of an array with 2 : 1 spacing ratios as used for measuring azimuth angles, but can be averaged out with a linear array.

incidence, and partly because the antenna beam is tilted slightly upwards.

A schematic diagram of the main elevation array is shown in *fig. 11a*. It has seven interferometer pairs, whose spacing increases by equal increments d. An iterative ambiguity-resolving process related to that described on p. 229 is used. If the unambiguous phase differences Φ_k thus derived are plotted as a function of the height h in multiples of d, the points will lie on a straight line through the origin (fig. 11*b*) in the absence of ground reflection. For the kth pair the phase difference is given by $\Phi_k = (2\pi kd/\lambda) \sin \beta$, and the slope of the line is $(2\pi/\lambda) \sin \beta$. This is therefore a measure of the elevation angle β.

procedures are used. One suitable scheme is to obtain four estimates of the slopes:

$$\frac{\Phi_7 - \Phi_3}{4d}, \quad \frac{\Phi_6 - \Phi_2}{4d}, \quad \frac{\Phi_5 - \Phi_1}{4d} \quad \text{and} \quad \frac{\Phi_4}{4d}.$$

An average slope σ_{av} can be derived from these values:

$$\sigma_{av} = \frac{1}{16d} (\Phi_7 + \Phi_6 + \Phi_5 + \Phi_4) - (\Phi_3 + \Phi_2 + \Phi_1), \tag{9}$$

and β can be derived from

$$\frac{2\pi}{\lambda} \sin \beta = \sigma_{av},$$

since d and λ are known.

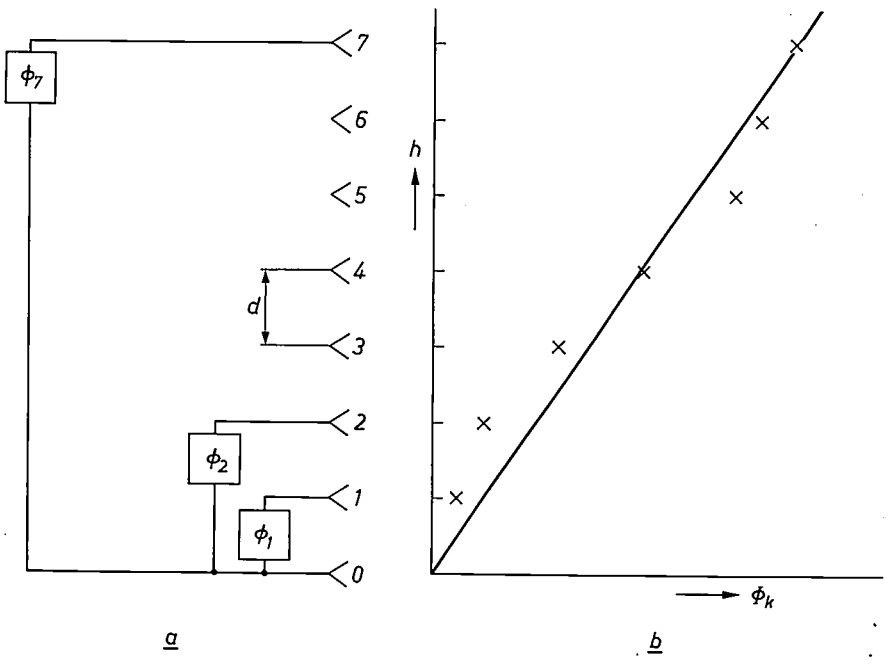


Fig. 11. *a*) Schematic diagram of the main elevation array. There are seven interferometer pairs, with the spacing increasing by equal increments d. *b*) With unambiguous phase differences Φ_1 to Φ_7 a plot of the phase differences against height (in multiples of d) will give a straight line. The slope $(2\pi d/\lambda) \sin \beta$ is a measure of the elevation angle β. (Note that in our diagram the axes are reversed from the conventional arrangement.) When there are ground reflections the points no longer lie on the straight line but are distributed around it as shown by the crosses. Nevertheless β can be determined by fitting the best straight line to the points. In practice these operations are performed by the logic units in the equipment.

Phase averaging

In the presence of ground reflections the seven points representing the phase differences Φ_k no longer lie on the straight line, but are distributed around it like the crosses in fig. 11*b*. Nevertheless the angle β can be determined by fitting the best straight line through the points. There are various formulae that can be used for fitting the straight line; since in our equipment the operations are performed by wired-logic units, simple

This method has the advantage of 'equal weighting' since an incorrect signal phase at any antenna provides the same error at the output. Amplitude nulls (and hence larger phase discontinuities) may therefore be allowed to appear at any position along the array. In practice, this means that the array may be operated at different heights above the ground, which may be necessary at different sites or with varying depths of snowfall.

Ambiguity resolution

Ambiguity resolution is based on an iterative approach related to equation (6) with the substitution $n = 1$. The phase differences for two successive interferometer pairs k and $k + 1$ in the array shown in fig. 11 are given by equations (3) and (4), but now they are taken to *include* the phase errors, i.e. they are the actual measured phase differences in digital form:

$$\Phi_k = \phi_k + 2\pi I_k, \quad (3)$$

$$\Phi_{k+1} = \phi_{k+1} + 2\pi I_{k+1}, \quad (4)$$

whence

$$\Phi_{k+1} = \Phi_k + \phi_{k+1} - \phi_k + 2\pi(I_{k+1} - I_k). \quad (10)$$

The quantities ϕ_k and ϕ_{k+1} are measured by the discriminators. I_{k+1} and I_k are unknown integer parts. Now provided that $|\Phi_{k+1} - \Phi_k| < \pi$ the quantity $I_{k+1} - I_k$ (the difference of the integer parts) can be determined from ϕ_k and ϕ_{k+1} in the way illustrated in *fig. 12*. This shows four possible arrangements of the phase differences, shown as vectors, from which it can be inferred that the following conditions apply.

Cases 1 and 2: $\phi_{k+1} - \phi_k < \pi$, $I_{k+1} - I_k = 0$, (11)

Case 3: $\phi_{k+1} - \phi_k < -\pi$, $I_{k+1} - I_k = +1$, (12)

Case 4: $\phi_{k+1} - \phi_k < +\pi$, $I_{k+1} - I_k = -1$. (13)

I_k and I_{k+1} are equal except when the two phase vectors are spread about the '12 o'clock position', in which case they differ by one. These conditions are also implicit in equation (6) with $n = 1$:

$$I_{k+1} = \text{nearest integer to } \frac{1}{2\pi}\left\{\phi_k + 2\pi I_k - \phi_{k+1}\right\}$$

$$\text{so that } I_{k+1} - I_k = \text{nearest integer to } \frac{1}{2\pi}\left\{\phi_k - \phi_{k+1}\right\}.$$

Thus Φ_{k+1} can be derived unambiguously from Φ_k, ϕ_{k+1} and ϕ_k, and starting with unambiguous measurement of Φ_1 from the closest-spaced pair, the other phase differences Φ_2, Φ_3 etc., can be derived unambiguously by the iterative process.

General arrangement

A block diagram of the elevation unit is shown in *fig. 13*. The spacing of the seven interferometers in the main array increases linearly in equal increments d of 18 cm, i.e. about 3λ. There is also a subsidiary array of three interferometers, with the spacing increasing linearly in equal increments d' of 9 cm. Horn antennas are used in both arrays; those in the main array have a horizontal beamwidth of 70°, a vertical beamwidth of 18° and are tilted upwards at an angle of 12°. The antennas in the subsidiary array have a horizontal beamwidth of 70°, a vertical beamwidth of 36° and are tilted upwards by 21°. The two arrays stand side by side with the separate common channels C and C' aligned at the same height at the bottom of the unit. As in the azimuth array the receivers are indicated by R and the discriminators by D. All the discriminators in the elevation unit digitize phase to 45°. The elevation unit also has 'wibwob' to allow measurement within the 45° quantization steps as explained before. This is applied to the common channel of the main array.

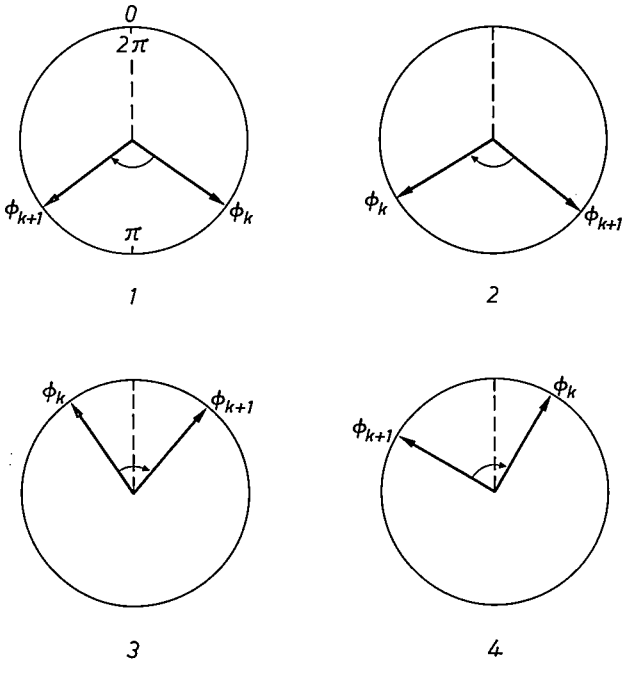


Fig. 12. Ambiguity resolution for the elevation array. Provided that $|\Phi_{k+1} - \Phi_k| < \pi$, the difference $I_{k+1} - I_k$ of the integer parts can be determined from ϕ_k and ϕ_{k+1}. The four possible arrangements of the phase differences are shown as vectors. The conditions (11) to (13) can be seen to apply. I_k and I_{k+1} are equal except when the two vectors are spread about the '12 o'clock position', when they differ by one. Thus Φ_{k+1} can be derived unambiguously from Φ_k, ϕ_{k+1} and ϕ_k, and starting with an unambiguous measurement of Φ_1 from the closest-spaced pair, the other phases can be derived unambiguously by the iterative process.

Processing logic

Similar kinds of processing logic are used in the two parts of the elevation array; we shall now describe the processing logic for the main array. Ambiguity resolution is carried out by a number of identical modules, one of which is shown schematically in the block diagram of *fig. 14*. The output signals from the discriminators k and $k + 1$ form 4-digit Johnson codes on parallel channels, and are each converted to pure

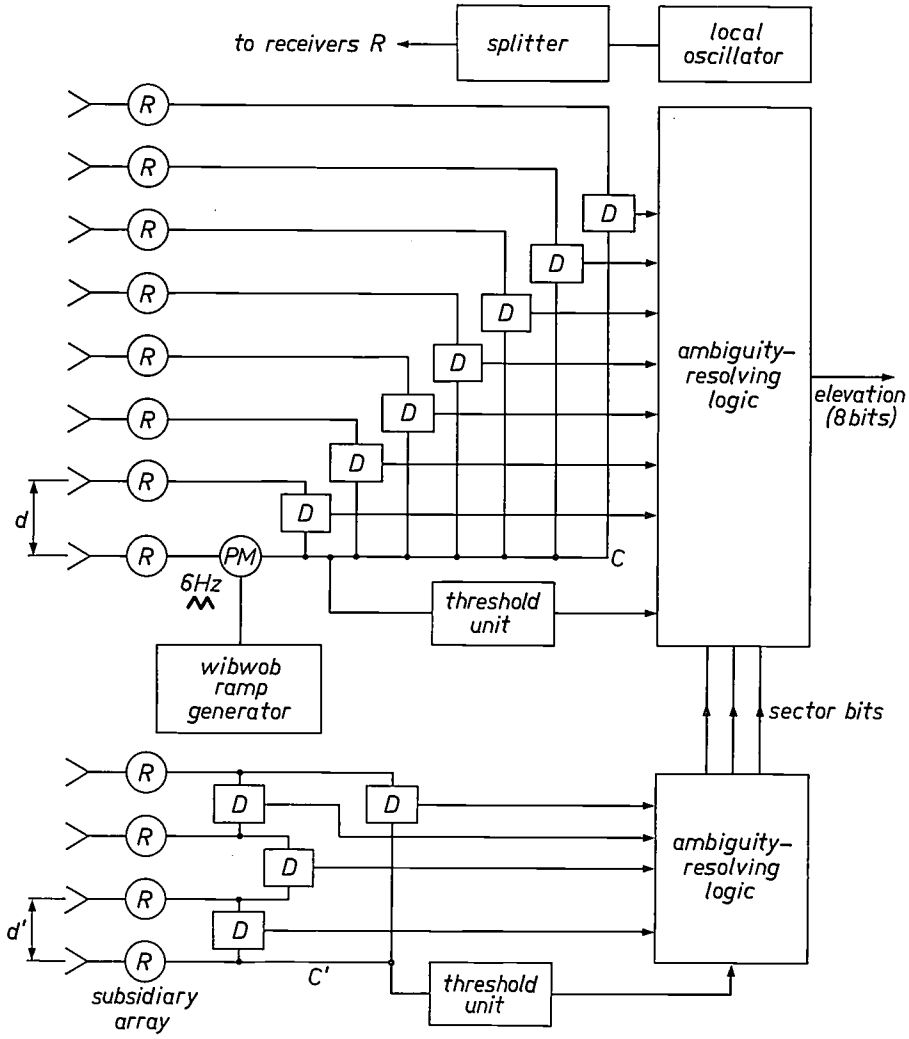


Fig. 13. Schematic diagram of the elevation unit. The main array has seven simple interferometers whose spacing increases linearly in equal increments d ($d = 18$ cm, i.e. about 3λ). There is also a subsidiary array of three interferometers with the spacing increasing in equal increments d' ($d' = 9$ cm). The two arrays stand side by side with the common channels C and C' at the same height. R receiver. D discriminator. PM phase modulator.

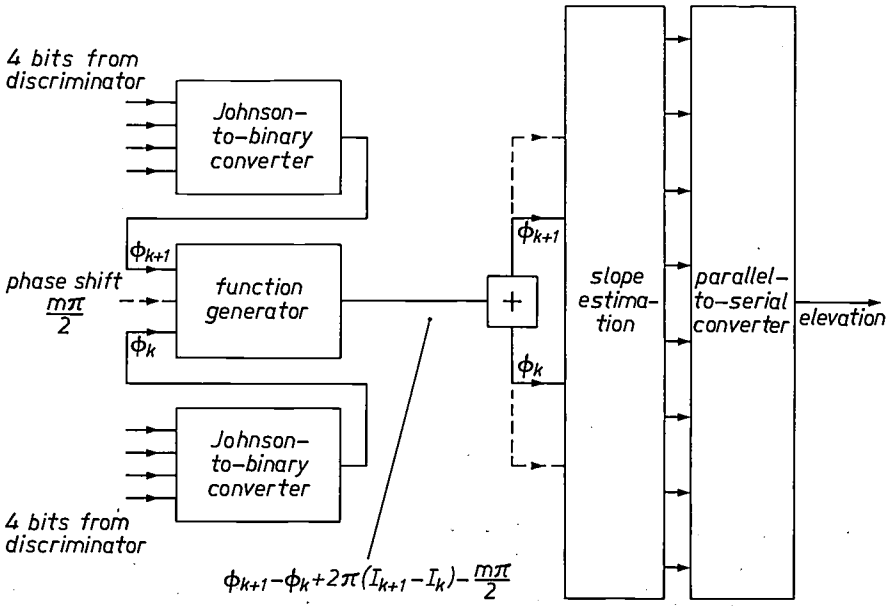
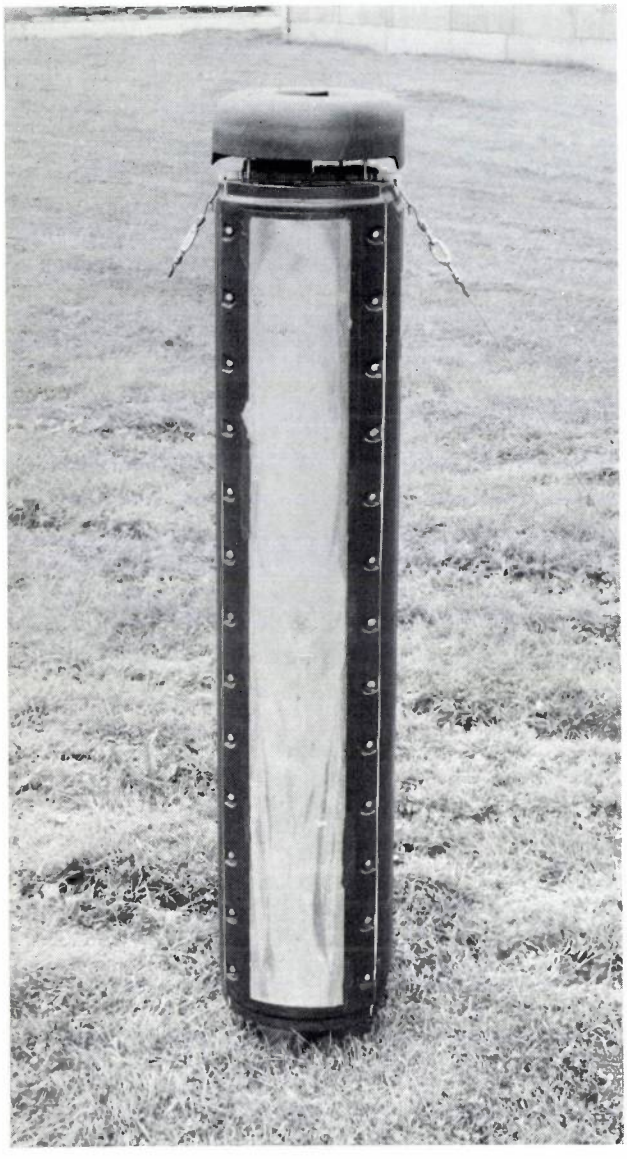


Fig. 14. Block diagram of processing logic for resolving ambiguous angle measurements from the elevation array. Output signals from the discriminators are converted to give pure-binary outputs proportional to ϕ_k and ϕ_{k+1}. The logic circuits perform the substitutions defined by conditions (11) to (13) and also the sum of eq. (10). This gives an unambiguous Φ_{k+1} from an unambiguous Φ_k, and hence an iterative ambiguity-resolution process. The unambiguous Φ values are arranged as in eq. (9) to produce an unambiguous binary output.



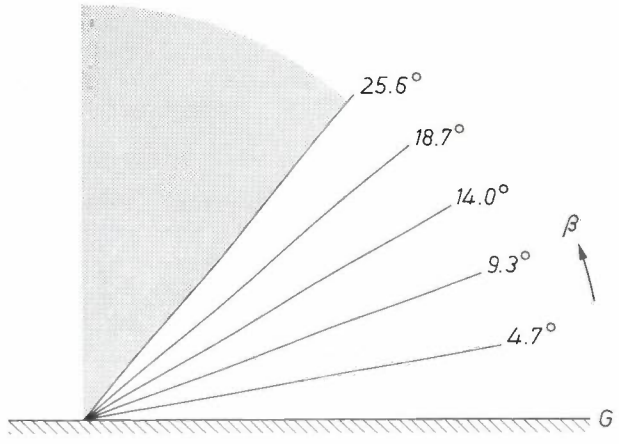
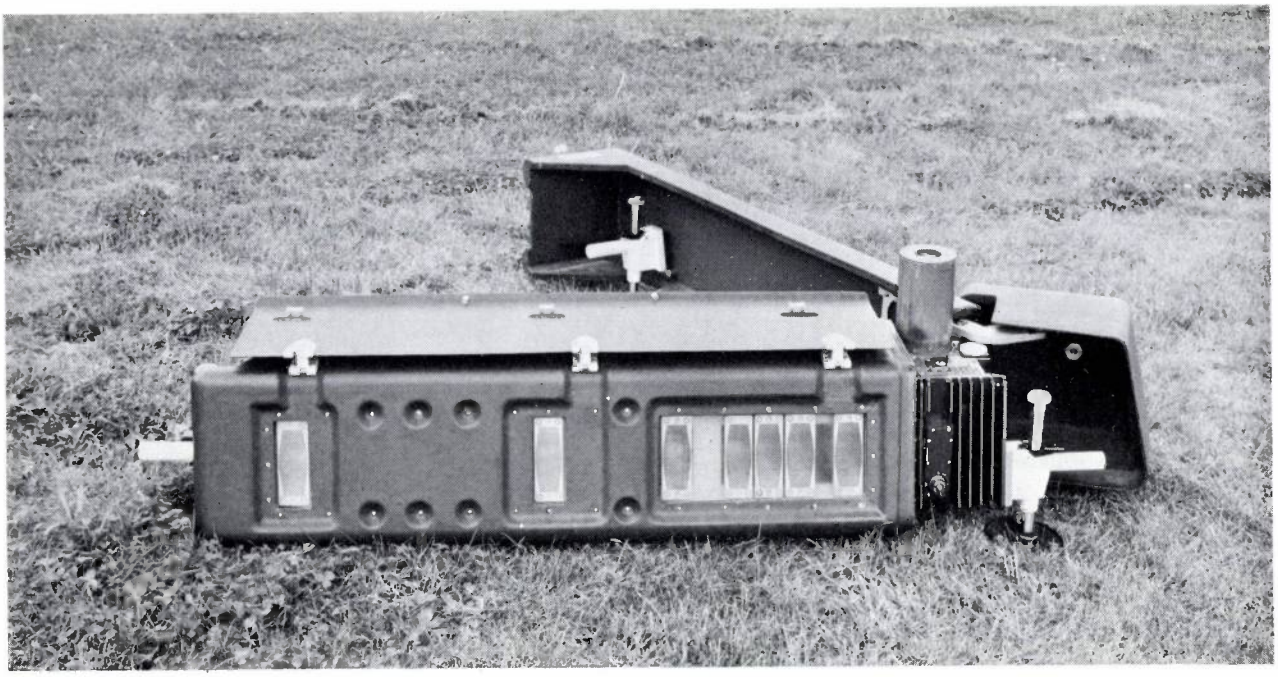


Fig. 15. Extension of elevation-angle coverage by the subsidiary array. *G* represents ground level. The angular coverage over which the logic circuits resolve the ambiguities by eq. (10) is limited to about 5° by the condition $|\Phi_{k+1} - \Phi_k| < \pi$ for the main array. The subsidiary array extends this basic coverage to 25° by producing angle-dependent phase steps that compensate the term $(2\pi d/\lambda) \sin \beta$ in $|\Phi_{k+1} - \Phi_k|$.

Fig. 16. The production-prototype interferometer units (without transponder). *Below:* azimuth unit. *Left:* elevation unit.



binary code. This provides two sets of parallel binary phase digits proportional to ϕ_k and ϕ_{k+1}. The logic circuits perform the substitutions defined by conditions (11), (12) and (13) and the sum defined by equation (10). When $m = 0$ (see fig. 14), the block labelled 'function generator' produces the quantity $\phi_{k+1} - \phi_k + 2\pi(I_{k+1} - I_k)$ at the output in binary code. The next block, the one labelled '+', adds to this number the quantity Φ_k. As eq. (10) shows, the result of this addition is Φ_{k+1}, so that we have an iterative ambiguity-resolution process starting with the unambiguous phase Φ_1. The unambiguous phase differences are then averaged as in eq. (9). The end result of the logic operations is a group of eight parallel binary digits, which are put into serial form for transmission over the data link to the aircraft

Extension of coverage by the subsidiary array

The angular coverage over which the logic circuits resolve the ambiguities by eq. (10) is limited to about 5° by the condition $|\Phi_{k+1} - \Phi_k| < \pi$ for the main array. The quantity $\Phi_{k+1} - \Phi_k$ contains errors arising in the phase measurement, errors due to ground reflections and an angle-dependent term $(2\pi d/\lambda) \sin \beta$.

The subsidiary array is used to extend this basic coverage to 25° by producing angle-dependent phase steps to compensate for the term $(2\pi d/\lambda) \sin \beta$. This smaller array operates in a similar way and gives an independent measurement of β. At each 5° step its logic circuits subtract a further $\pi/2$ from the difference $|\Phi_{k+1} - \Phi_k|$. (In sector m, where m is an integer between 0 and 4, a phase difference $m\pi/2$ is subtracted from the function $(\phi_{k+1} - \phi_k) + 2\pi(I_{k+1} - I_k)$.) In this way $\Phi_{k+1} - \Phi_k$ is kept well below the value π, so that additional errors do not affect ambiguity resolution, and the accuracy of the main array is available for five ranges of about 5°, giving a total coverage of 25.6° (see *fig. 15*).

The prototype equipment

Photographs of production-prototype interferometer units are shown in *fig. 16*. Light-weight construction methods are used to permit easy handling. The units are equipped with sights and levelling bubbles for setting up. Legs, feet and other protrusions can be folded away for transit, and the cases are designed to withstand vibration and impact.

Extensive flight tests of the MADGE prototype equipment have been made at the Royal Aircraft Establishment, Bedford, England, over a total of about 500 flying hours. Comparisons were made of the aircraft's position as indicated by the interferometers and the position indicated by an accurate optical tracking system based on a kine-theodolite. (This is a kind of cine camera that records on each frame a very accurate record of the direction in which the camera is pointing. Its accuracy is about 0.0001°.)

Fig. 17 shows three different types of flight path used in the tests. In the example of fig. 17*a* the aircraft describes roughly semicircular paths to investigate the measurement of the azimuth angle θ (approach and overshoot units are similar). A typical result is given

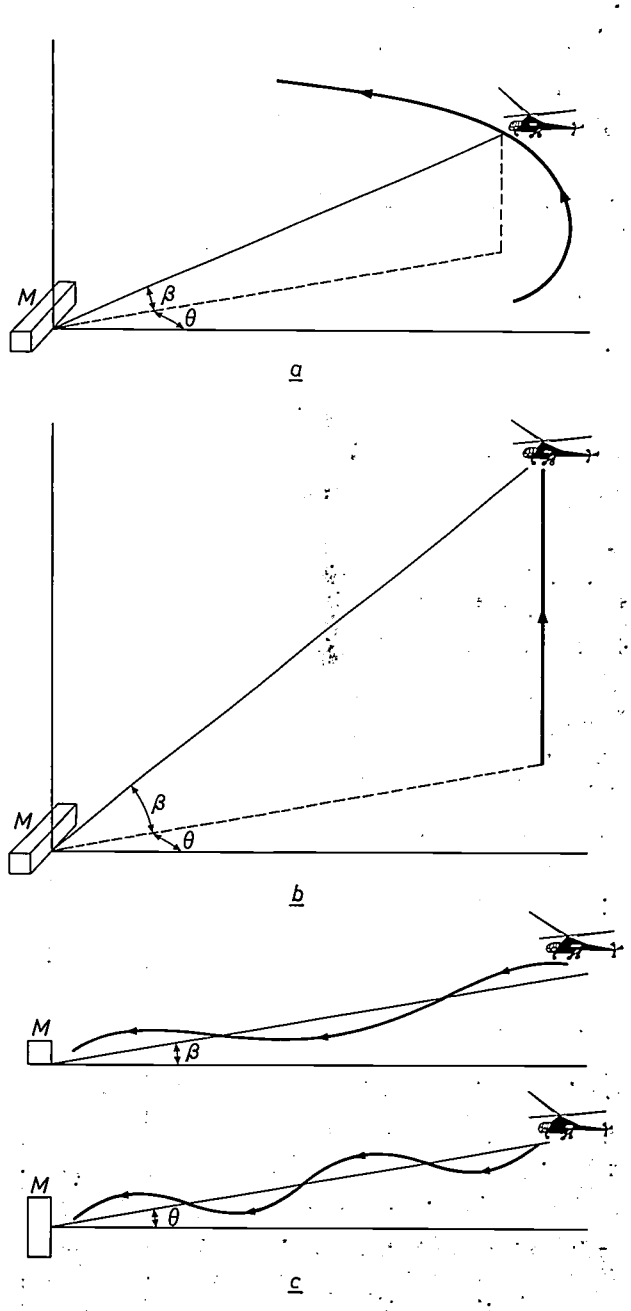


Fig. 17. Three different types of flight path used in tests on MADGE. The box M represents the MADGE equipment. *a*) Roughly semicircular path at constant height (semi-orbit), to investigate the measurement of azimuth angle θ. *b*) Vertical rise, to investigate the measurement of elevation angle β. *c*) Paths flown to provide a more direct comparison with the performance of an ILS system. The upper diagram is an elevation, the lower a plan.

in *fig. 18*, which shows that the coverage is $\pm 65°$ for the azimuth angle θ. The vertical rises in fig. 17*b* were used for a corresponding investigation of the measurement of elevation angle β; a typical result is shown in *fig. 19*. The curve is essentially linear for elevation angles β down to about 1°. Fig. 17*c* shows the approach paths that were flown to provide a more direct comparison with the performance of an ILS system,

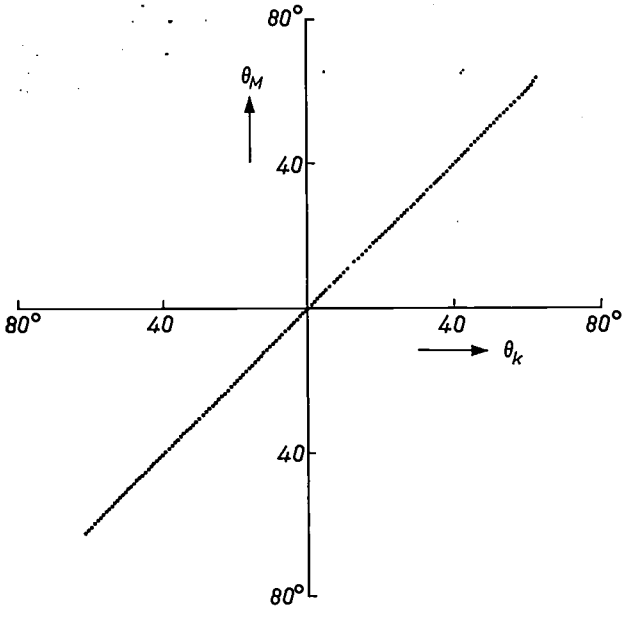


Fig. 18. Azimuth-angle characteristic derived from results of measurements made with a flight path as in fig. 17*a*. θ_k azimuth angle indicated by kinetheodolite. θ_M azimuth angle indicated by MADGE. The coverage is $\pm$ 65°.

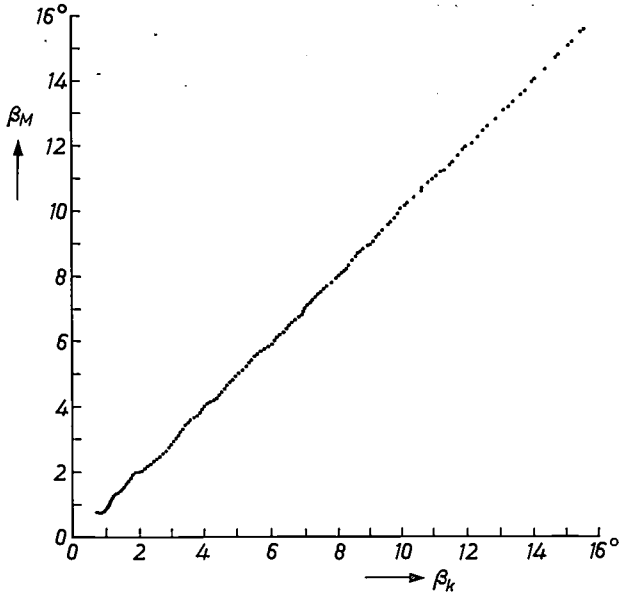


Fig. 19. Elevation-angle characteristic derived from results of measurements made with a flight path as in fig. 17*b*. β_k elevation angle indicated by kinetheodolite. β_M elevation angle indicated by MADGE. The curve is essentially linear down to about $\beta_k = 1°$.

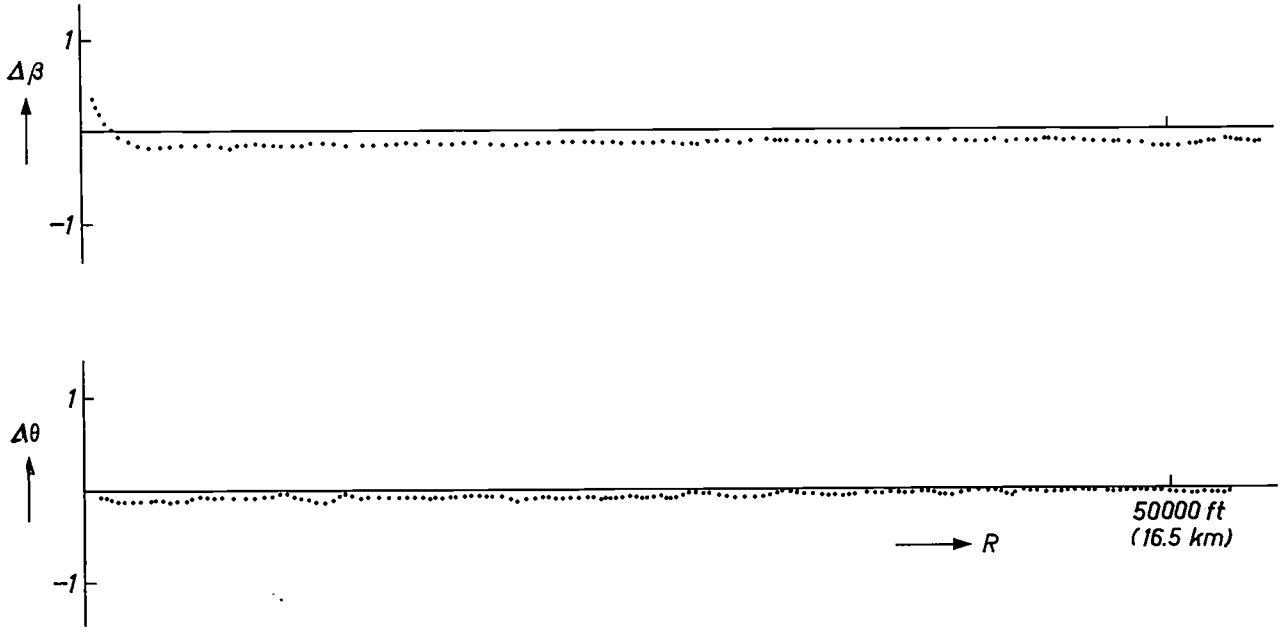


Fig. 20. Typical results with the flight path of fig. 17*c*, based on 3° glide slope. $\Delta\beta$ elevation error against kinetheodolite. $\Delta\theta$ azimuth error against kinetheodolite. R range. These curves also give the systematic errors of the measurement system.

which is always designed to work with a single path. *Fig. 20* shows some typical results. The performance figures shown earlier in the article in Table I were derived from the results of these tests.

Summary. MADGE is a microwave landing-guidance aid providing accurate guidance over wide angles of coverage and distances up to 27 km. It has passive interferometer arrays for measurement of azimuth and elevation angles and a data link that supplies guidance information to the aircraft. Approach angles can be selected by the pilot; horizontal and vertical position errors and range are displayed in the aircraft. The system is capable of providing simultaneous guidance for over 150 aircraft. The ground equipment is battery operated and portable; two men can set it up in 15 minutes at temporary landing sites.

A short general treatment of measurement of angle with multiple interferometers, including ambiguity resolution and possible phase errors, is followed by descriptions of the phase-measuring receivers and the azimuth and elevation units. The interferometer spacings in the main azimuth array are arranged in a geometric progression, but those of the elevation array are arranged in an arithmetic progression. Ambiguity resolution and processing logic are different for azimuth and elevation arrays. Flight tests have shown that the angular accuracy (about 0.05°) is equivalent to that of a category II ILS system over a wide region of coverage: azimuth angles are indicated over a 90° azimuth sector for an elevation range of 1-40°; elevation angles are indicated over 1.5-25° for a horizontal sector of 90°.

Philips tech. Rev. 34, 242-243, 1974, No. 9

Double-glazed windows with very good thermal insulation

In buildings heat losses through windows generally constitute a substantial part of the total energy consumption. This applies even when double glazing is used, because its thermal conductivity k is still a multiple of that of a well insulating wall. In the best insulated outside walls k values of the order of 0.6 W/m²K can nowadays be achieved, and the rising cost of energy provides a strong incentive to achieve even lower values. The heat loss through windows would then become relatively even greater — unless one were to make do with smaller windows. There is thus an evident need for windows whose thermal conductivity is substantially lower than that of the double-glazed windows in current use ($k = 3.15$ W/m²K).

This need can be met by installing windows like those illustrated in *fig. 1*. A window of this type differs in two respects from conventional double-glazed windows. Firstly, each pane is coated on the inside with a thin layer of material that transmits visible light but strongly reflects infrared [1]. Secondly, the space between the panes is filled not with air but with a gas that has a lower thermal conductivity. These two measures reduce the value of k very considerably, to a value calculated at about 0.9 W/m²K.

The curve in fig. 1 shows a temperature profile that could be encountered in a particular situation. In spite of the temperature difference of 23 °C between the inside and the outside, the heat flux through the window is only 25 W/m²; in the same situation the loss would be 70 W/m² through a normal double-glazed window. The considerable saving of energy made possible by the new windows suggests that they will already be an economic proposition in spite of their higher price.

The physical explanation of the improvement obtained is as follows. The outside and inside panes acquire approximately the temperature of the outside and inside air respectively, since the air is always more or less in movement. The actual heat resistance of the window is located in the area between the two panes. The distance between them is small enough to suppress convection in this space, but large enough for a sufficiently low thermal conductivity. The two glass surfaces facing one another exchange heat not only by conduction through the filler gas but also by radiation. In a conventional double-glazed window, with an air cavity 12 mm wide, the contribution of radiation to the heat transport is roughly 2/3, and that of conduction only about 1/3.

The part k_s of k that defines the radiation contribution in double-glazed windows is given by the expression:

$$k_s = 4\sigma T^3 \frac{\varepsilon^2}{1-(1-\varepsilon)^2},$$

where σ is the Stephan-Boltzmann constant, T the mean absolute temperature of the two panes, and ε the thermal emissivity of the surfaces facing each other. In the wavelength region concerned (5-50 μm), ε for glass is about 0.9. However, the value of ε for glass coated with a layer of doped SnO_2 or In_2O_3 is no greater than 0.1 to 0.2. The heat transmission by radiation in windows made of such glass is therefore about ten times smaller than in other windows. It is obviously useful to fill the space between such windows with a gas that has a lower thermal conductivity than air. The heavy inert gases, such as krypton, can be used for this purpose, but some other gases of high molecular weight can also be used.

A detailed picture of how the infrared-reflecting layers work is given in *fig. 2*, where the reflection R and the transmission D of a tin-doped In_2O_3 layer are plot-

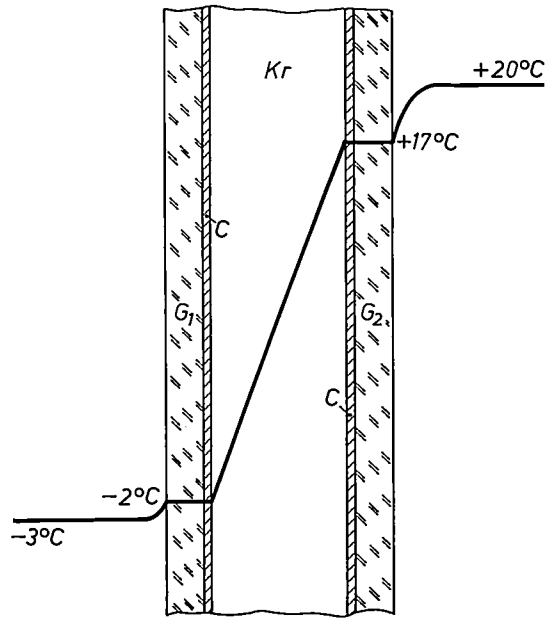


Fig. 1. Cross-section of a window with very high thermal insulation. G_1 outside pane. G_2 inside pane. C layer of doped SnO_2 or In_2O_3. The space between the two panes is filled with a gas of low thermal conductivity. The curve gives the temperature profile of a window coated with a layer of SnO_2 and using a krypton gas filling, for the case where the outside temperature is −3 °C and the inside temperature +20 °C. The heat flux density through the window in this case is only 25 W/m², compared with 70 W/m² for a conventional double-glazed window.

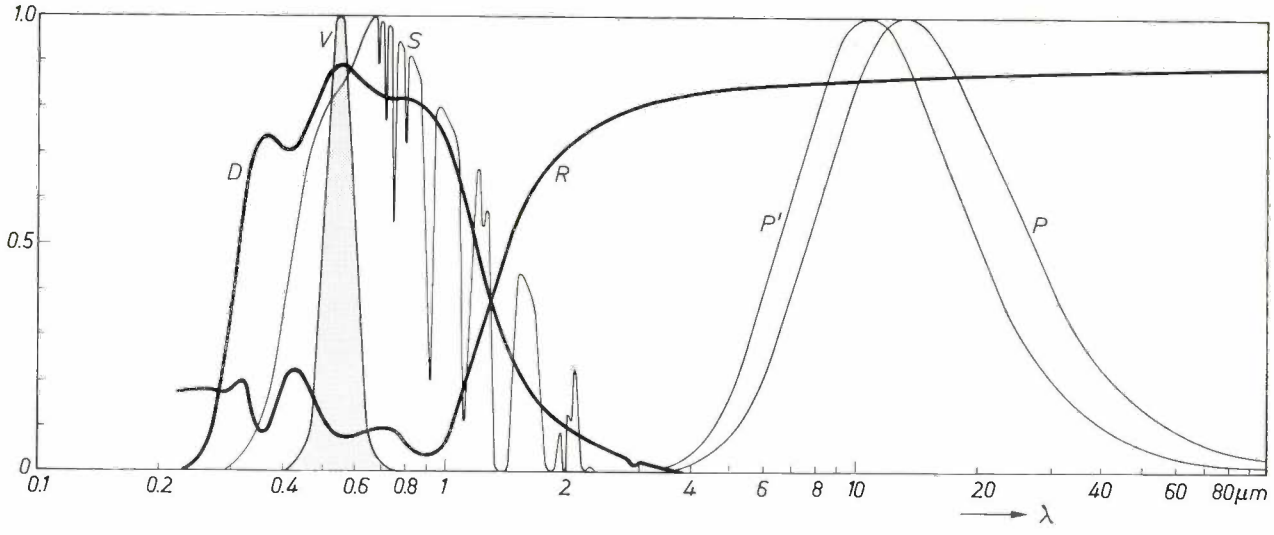


Fig. 2. Transmission D and reflection R of an optimally tin-doped In_2O_3 layer 0.15 μm thick, as a function of the wavelength λ. Also given, for comparison, are the spectral luminous-efficiency characteristic (curve V), the spectral energy distribution of the Sun (curve S), that of a black body at 273 K (curve P) and its derivative to absolute temperature (curve P'; this quantity is a measure of the energy transfer by radiation between black surfaces that differ relatively little in temperature). In the visible region the transmission is almost equal to that of glass; in the infrared the reflection is about 0.85, and the emission factor ε is therefore only about 0.15.

ted as functions of the wavelength. Comparison with the other curves shown in the figure — the spectral luminous-efficiency curve for the human eye, the spectral energy distribution of the Sun, that of a black body at 273 K and its temperature derivative — show how very well the properties of the layer meet the requirements. The strong infrared reflection is due to free electrons from the tin dope. The concentration of these electrons is at a maximum (about 1.5×10^{21} cm^{-3}) when the dope concentration is about 10 at%. The plasma wavelength, which determines the significant change of the optical characteristics, then lies at about 1.1 μm. The maxima and minima in the curves in the region below 1 μm are due to interference. Given a layer of optimum thickness it is possible to obtain maximum transmission in the region of greatest spectral luminous efficiency. The value of D is then very little lower than that of ordinary glass.

H. Köstlin

[1] Similar layers were introduced some years ago in Philips low-pressure sodium lamps (type SOX). See for example H. J. J. van Boort and R. Groth, Philips tech. Rev. **29**, 17, 1968.

Dr H. Köstlin is with Philips Forschungslaboratorium Aachen GmbH, Aachen, West Germany.

Philips tech. Rev. 34, 244-254, 1974, No. 9

Vacancy clusters in dislocation-free silicon and germanium

A. J. R. de Kock

The beauty and symmetry of crystals in their outward forms has long deceived us into attributing to them an internal perfection that is illusory. Since about 1930, when the study of lattice defects can be said to have begun, crystallographers have built up a minutely detailed picture of the many ways in which the internal structure of crystals may deviate from geometrical perfection. In recent years, however, the semiconductor industry has demanded a higher and higher internal perfection in certain crystals. Methods were devised for growing crystals of germanium and silicon free of dislocations, and these crystals were used in fabricating improved devices. Later, the very absence of these defects gave rise to troublesome effects in certain cases from the formation of vacancy clusters. How these microdefects have been dealt with is outlined in the article below.

Silicon and germanium single crystals are used on a vast scale as the staple materials of the semiconductor industry. The use of germanium is nowadays largely restricted to diodes and transistors made by alloying and diffusion techniques. Silicon, on the other hand, is almost exclusively the material used for planar devices and integrated circuits. Silicon and germanium crystals also find application in special devices such as lithium-drifted radiation detectors. A potential application of silicon is in vidicon-type camera tubes.

The electrical behaviour of all these devices is very much dependent on the purity, homogeneity and perfection of the crystals used. Techniques were developed some twenty years ago whereby single crystals can be grown entirely free of dislocations [1]. Furthermore, the floating-zone technique of crystal growth [2] has long been used to yield silicon crystals of such a high purity that impurity precipitates are also absent.

Nevertheless, these crystals are not absolutely perfect. Especially in the absence of dislocations and precipitates certain types of microdefects are formed that have adverse effects on the electrical performance of semiconductor devices [3]. We have been able to establish that these microdefects are vacancy clusters of various types [4]. They are formed in the growing crystal behind the interface between solid and liquid and are distributed over the crystal in such a way that patterns of spiral striations ('swirls') are produced. *Fig. 1* shows an example.

Dr Ir A. J. R. de Kock is with Philips Research Laboratories, Eindhoven.

This article is concerned particularly with these vacancy clusters: their nature and properties, how they are formed and how they may be detected. Finally it is shown how they may be eliminated. The main emphasis will be on silicon.

Point defects in melt-grown crystals

Point defects in crystals grown from the melt may be classed into two groups — chemical impurities, both wanted and unwanted, and thermally generated point defects such as interstitials and vacancies. We shall first look at the various point defects in silicon and germanium to see if we can decide which defect is responsible for the swirls in silicon and germanium crystals.

Chemical impurities

Under clean conditions silicon and germanium crystals can be grown in which the concentration of metallic impurities is well below the limits of detection by chemical methods, mass spectrometry or neutron-activation analysis ($< 10^{13}$ cm^{-3}). These low concentrations can be achieved primarily because of the purity of the polycrystalline starting material and the low value of the distribution coefficients [2] of metals in silicon and germanium, i.e. the ratio of the impurity concentration in the solid and liquid states.

The impurity carbon, occupying substitutional lattice positions in silicon, is frequently present in larger concentrations. In crystals grown by Czochralski's method from the melt in a crucible the concentration

Fig. 1. Monitor picture from a silicon vidicon[3] with a target of dislocation-free silicon (cross-section 25 mm) operating in darkness. The bright spots are leaky diodes; they are distributed in a spiral pattern. The continuous spiral pattern is background leakage, probably due to impurities. Magnification 9×.

may be as high as 5×10^{17} cm^{-3}. In crystals grown by the floating-zone technique however, under the best conditions, the carbon concentration may be below 2×10^{15} cm^{-3}.

The oxygen concentration in silicon depends on the crystal-growth conditions such as the oxygen concentration in the polycrystalline starting material and the partial pressure of oxygen in the atmosphere used. Growth under high-vacuum favours a low oxygen content; not only is the partial pressure of oxygen then low but evaporation of oxygen in the form of SiO is favoured. The rate of rotation of the crystal during growth also affects the oxygen content. Fast rotation causes intensive stirring of the melt and in the Czochralski method of growth this increases the rate at which crucible material (silica) dissolves. The oxygen content of Czochralski-grown crystals (10^{17}-2×10^{18} cm^{-3}) is therefore generally higher than that of floating-zone crystals (10^{14}-10^{16} cm^{-3}). *Table I* summarizes the impurity concentrations usually found in melt-grown silicon crystals.

Since the solubility of impurities in silicon and germanium decreases as the temperature decreases the crystals become supersaturated with oxygen and carbon on cooling. Supersaturation with metallic impurities can hardly occur because of their low concentrations. In floating-zone silicon supersaturation only begins at about 300 °C with oxygen and 800 °C with carbon. Since the diffusion coefficient at this temperature is very low, precipitation does not occur. For Czochralski-grown Si crystals, however, which become supersaturated at higher temperatures, precipitation can be expected to occur. Germanium crystals are nearly always grown by means of the Czochralski method. Nevertheless the oxygen and carbon concentration remain relatively low since germanium does not react with the material of the crucible at its melting point (936 °C).

Table I. Typical concentrations (atoms per cm^3) of oxygen, carbon and metals in silicon crystals made by the Czochralski and floating-zone methods.

	Oxygen	Carbon	Metals
Czochralski	10^{17}-2×10^{18}	4×10^{16}-5×10^{17}	$<5\times10^{13}$
Floating-zone	10^{14}-10^{16}	2×10^{15}-10^{17}	$<5\times10^{13}$

Apart from these unwanted impurities the crystals also contain impurities deliberately added to the material (dopants), such as phosphorus, arsenic, antimony (donors, giving free electrons and thus *N*-type material) or elements such as boron or indium (acceptors, yielding excess holes and thus *P*-type material). In practice, doping levels are generally kept so low that the material contains no precipitates of these impurities.

A general conclusion from the foregoing is that some carbon and oxygen precipitation is possible in Czochralski-grown silicon crystals but that in germanium and in floating-zone silicon crystals impurity precipitation is negligible.

Impurity striations

As mentioned earlier, crystals drawn from the melt are rotated during growth. Although this reduces thermal asymmetries, a perfectly cylindrical-symmetric temperature distribution around the axis of rotation is generally not achieved. Consequently, every point on the solid-liquid interface will become alternately hotter and cooler during rotation. This results in a periodic

[1] W. C. Dash, in: R. H. Doremus, B. W. Roberts and D. Turnbull (ed.), Growth and perfection of crystals, Wiley, New York 1958, p. 361. See also B. Okkerse, Philips tech. Rev. **21**, 340, 1959/60.

[2] A survey of zone-melting methods can be found in J. Goorissen, Philips tech. Rev. **21**, 185, 1959/60.

[3] An account of the silicon vidicon is given in M. H. Crowell and E. F. Labuda, Bell Syst. tech. J. **48**, 1481, 1969.

[4] A. J. R. de Kock, thesis, Nijmegen 1973 (also published as Philips Res. Repts. Suppl. 1973, No. 1). A first experimental indication of the existence of microdefects was found as early as 1965 by T. S. Plaskett (Trans. Met. Soc. AIME **233**, 809, 1965), who also suspected that vacancy clusters were involved.

variation in growth rate. Because of this the impurity concentrations will in general exhibit a periodic variation along the crystal-growth direction.

The fastest-growing regions of the crystal, where the impurity concentration will be highest, form a spiral ramp with a pitch equal to the pulling rate divided by the rotation rate. In a longitudinal crystal section parallel to the crystal axis the alternating regions of high and low impurity concentration will form a system of narrow bands nearly parallel to the solid-liquid interface [5]. These bands are called impurity striations. When the impurity concentration is high as, for instance, in heavily doped material, these striations can be revealed by means of several techniques such as preferential etching, electrolytic copper plating and X-ray topography. A *cross*-section of the crystal then shows a spiral pattern.

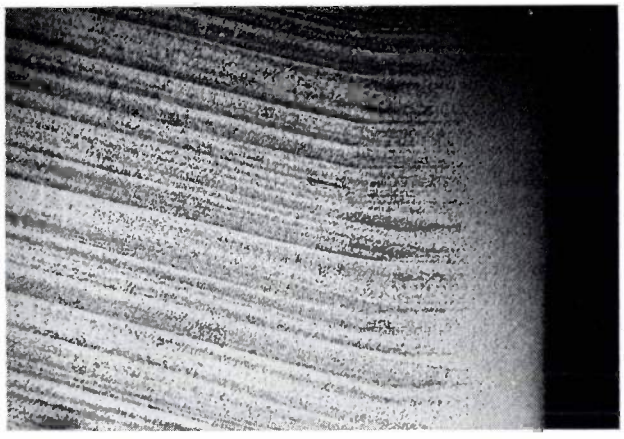


Fig. 2. Preferentially etched longitudinal section of the edge of a dislocation-free silicon crystal grown by the floating-zone technique in the ⟨111⟩ direction. Magnification 16×. The distribution of the etch pits in striations can be clearly seen. The strip next to the outer surface of the crystal contains no etch pits.

The relation between the crystal-growth rate and the impurity follows directly from the expression for the effective distribution coefficient [6]:

$$K_{\text{eff}} = \frac{K}{K + (1 - K)\exp(-V_g\delta/D)}.$$

Here K is the equilibrium distribution coefficient as determined by the phase diagram for the binary system silicon/impurity, V_g is the crystal-growth rate, δ the thickness of a layer in the melt adjacent to the solid-liquid interface where transport of impurities takes place solely by diffusion (δ is about 10^{-2} cm), and D the diffusion coefficient of the impurity in the melt (10^{-4}-10^{-5} cm²/s). The values of K quoted in the literature are smaller than unity for all important impurities in silicon. Taking, for instance, K equal to 0.1, $D = 10^{-4}$ cm²/s and a growth rate $V_g = 5\times10^{-3}$ cm/s, as is frequently used, we obtain: $K_{\text{eff}} = 1.6\,K = 0.16$. This is 1.6 times the equilibrium value (K) that would be expected for slow growth rates.

Vacancies and interstitials

Vacancies and interstitials are thermodynamically stable at elevated temperatures because their formation lowers the free energy of the crystal. Although considerable energy (enthalpy) is required for the formation of these point defects, this is compensated at relatively high temperatures by the increase in entropy. The energies of formation of both types of point defect, however, are not accurately known. Theoretical estimates for silicon yield values between 2.3 and 4.6 eV for the formation energy of a vacancy. The formation energy of an interstitial is even less accurately known but it is assumed to be higher.

The equilibrium concentration N_e for both kinds of defect will be governed by Boltzmann statistics:

$$N_e = N\exp(-\Delta G/kT) = N\exp(\Delta S/k)\exp(-E/kT), \quad (1)$$

where N is the number of atomic or interstitial sites, ΔG is the free energy of formation, E is the formation energy and ΔS is the entropy of formation. Exact values for ΔG are not known. It is well known, however, that the equilibrium concentration of vacancies in metals with face-centred cubic (f.c.c.) and body-centred cubic (b.c.c.) lattices is much higher than that of interstitials; the formation entropies in these metals are approximately equal but the formation energies for the interstitials are much higher. This difference is connected with the lattice strain introduced by the interstitials. The introduction of vacancies is accompanied by only very small strain effects. A similar argument can be applied for silicon. It appears that vacancies are therefore the more important of the thermally generated point defects.

From thermodynamical considerations A. Seeger and M. L. Swanson [7] estimated the upper limit for the equilibrium concentration of vacancies in silicon at the melting point to be 9×10^{15} cm^{-3}, and the lower limit 1.2×10^{13} cm^{-3}. Values obtained from quenching experiments, however, seem to indicate concentrations much higher than 10^{16}. For germanium, the vacancy concentration is estimated to be about 3×10^{15} cm^{-3}.

Equation (1) shows that the equilibrium concentration for vacancies decreases with decreasing temperature. Silicon or germanium crystals grown from the melt therefore become supersaturated with vacancies during cooling. Dislocations act as sinks for vacancies and, because vacancies have a high diffusion coefficient in germanium and silicon, even a small concentration of dislocations is sufficient to clear the whole crystal of its excess vacancies. In material with no dislocations, however, the high supersaturation can only be eliminated by vacancy *precipitation*. This leads to the formation of the vacancy clusters mentioned in the introduction. In the next section we shall briefly describe the detection methods we have used to make these vacancy clusters visible.

Detection of microdefects in silicon and germanium

A simple method developed for revealing dislocations and stacking faults is preferential etching, for example with the etchant developed by E. Sirtl and A. Adler [8], a solution of CrO_3 and HF in water. Because of the extreme sensitivity of this etchant to small local strain fields, it can also be used to reveal microdefects. *Fig. 2* shows a preferentially etched longitudinal section of a dislocation-free silicon crystal grown by the floating-zone technique. Etching of cross-sectional crystal slices reveals shallow etch pits in a pattern of interrupted rings.

Another method that has long been known for revealing imperfections in translucent crystals is to 'decorate' them with precipitates of impurities. This has been mainly used for revealing dislocations, for example in silver halides, alkali halides, germanium and silicon. Since germanium and silicon do not transmit visible light but do transmit infrared, an infrared microscope is used in these cases. Microdefects such as vacancy clusters can be decorated as well as dislocations; in silicon this can be done with copper. In our method copper is diffused into silicon samples by heating them at 950 °C in a stream of argon in which copper is evaporated. After saturation the sample is quenched to room temperature, resulting in precipitation of copper on the lattice defects present (*fig. 3*).

The copper-decoration technique cannot be used for germanium, for one reason because copper is not very soluble in germanium. There are also difficulties in using the method with silicon. During diffusion at 950 °C, new microdefects may be generated and other microdefects that *were* present but are unstable at high temperatures may dissociate and remain undetected. The lithium-decoration technique that we have developed, applicable to both silicon and germanium, is an improvement in these respects since the required lithium diffusion takes place at much lower temperatures. Lithium decoration yields information complementary to that given by copper decoration.

Lithium decoration

Lithium atoms in silicon and germanium take up interstitial positions in the lattice, since they are very small. For these reasons the diffusion coefficient of lithium is high even at relatively low temperatures. In addition, the solubility of lithium in silicon and germanium is about 1000 times greater than that of copper in silicon at the same temperature. High concentrations of lithium can therefore be introduced by diffusion at quite low temperatures (e.g. in Si 8×10^{17} cm^{-3} at 400 °C) in a reasonably short time (some hours). As with copper, cooling causes supersaturation, resulting in precipitation. This takes place on various types of

Fig. 3. Infrared micrograph of a region at the edge of a silicon crystal grown in the ⟨111⟩ direction. The microdefects have been made visible by 'decorating' with copper. Magnification 170×.

nucleation sites, e.g. stacking faults and dislocations and, as we shall see shortly, on one particular type of vacancy cluster.

The cooling may be carried out slowly, resulting in Li precipitation during the cooling procedure itself. However, in the case of quenching to room temperature, a subsequent anneal at about 160 °C for Si and 40 °C for Ge is required for a sufficiently strong decoration to take place. The Li-decorated defects have been studied by infrared microscopy (*fig. 4*) and in particular with the X-ray transmission topographical technique (see below).

X-ray transmission topography

Several techniques of X-ray topography are available for the investigation of imperfections in nearly perfect crystals. The topographs can be made either in reflection (Bragg geometry) or in transmission (Laue geometry). Imperfections such as dislocations or precipitates are recorded on these topographs because the lattice distortion around the defects influences the diffracted X-ray intensity.

We have used the transmission scanning method due to A. R. Lang [9]. *Fig. 5* is an example of a transmission

[5] J. A. M. Dikhoff, Philips tech. Rev. **25**, 195, 1963/64.
[6] J. A. Burton, R. C. Prim and W. P. Slichter, J. chem. Phys. **21**, 1987, 1953.
[7] A. Seeger and M. L. Swanson, in: R. R. Hasiguti (ed.), Lattice defects in semiconductors, Univ. of Tokyo Press, 1968, p. 93.
[8] E. Sirtl and A. Adler, Z. Metallk. **52**, 529, 1961.
[9] A description of this method can be found in A. E. Jenkinson, Philips tech. Rev. **23**, 82, 1961/62.

Fig. 4. Infrared transmission micrograph of a dislocation-free silicon crystal, grown in vacuum, in which lithium has been diffused. The round patches of precipitated lithium have formed at vacancy clusters. The lithium diffusion was carried out at 600 °C and was followed by slow cooling and annealing at 160 °C for 120 hours. Magnification 170×.

X-ray topograph of a sample exhibiting 'direct-image' contrast. The photograph shows a cross-sectional slice cut from a silicon crystal with a moderate dislocation density.

However, the microdefects, which we are particularly interested in, cannot be seen in this way. The associated lattice distortions extend over such a small field (less than about 10^{-4} cm) that the diffracted intensity originating from these microdefects is masked by the

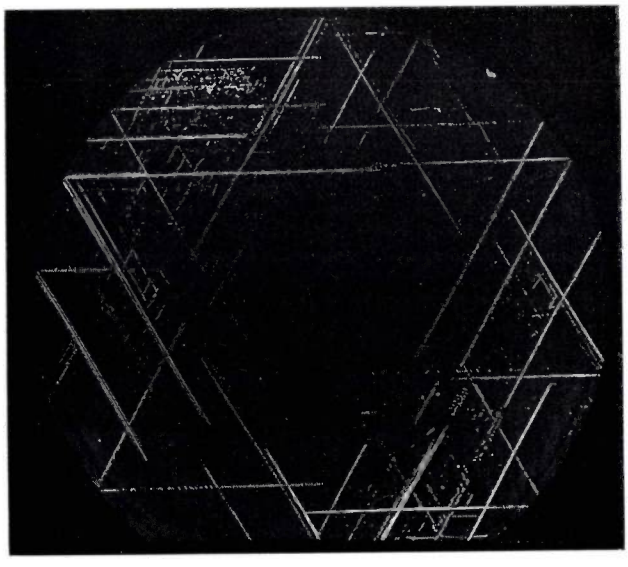


Fig. 5. X-ray transmission topograph of a silicon slice (diameter 24 mm). The dislocations in the crystal have been made visible.

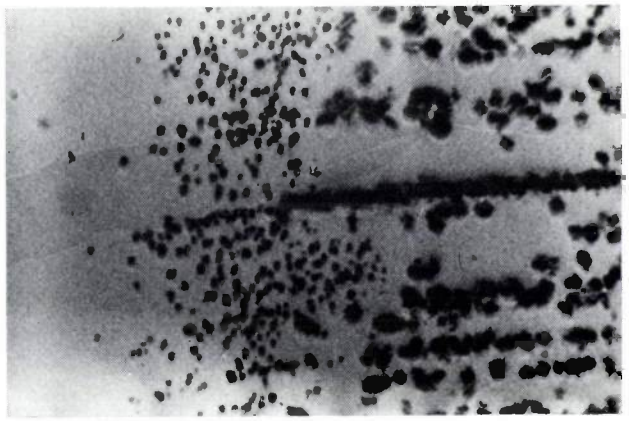


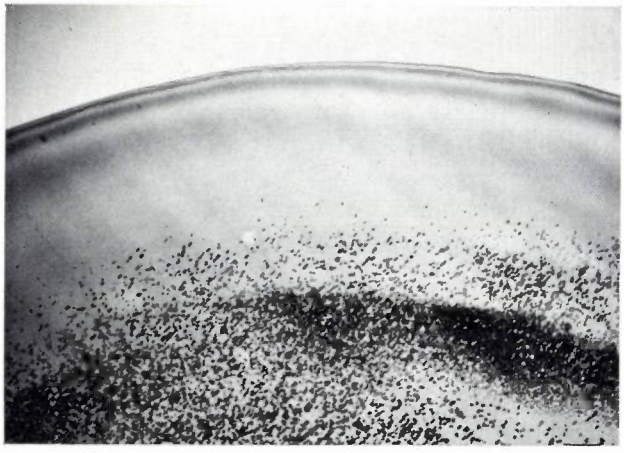
Fig. 6. X-ray topograph of the same region of the copper-decorated crystal as shown in fig. 3, decorated with copper (about the same magnification). The same defects have been made visible.

background intensity diffracted by the perfect regions of the crystal. However, the lattice distortion can be enlarged artificially by decorating the microdefects with copper or with lithium. The strain field around the decorated defects can be made sufficiently large to cause strong direct-image contrast. This combination of decoration and X-ray topography has been found to be of paramount importance for the study of vacancy clusters. *Figs. 6* and *7* show X-ray topographs of silicon crystals decorated with copper and lithium respectively.

Character of the microdefects in dislocation-free crystals

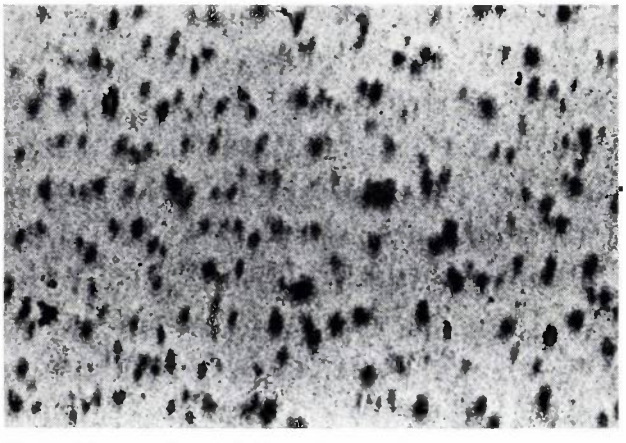
Cross-sectional and longitudinal slices of various floating-zone dislocation-free crystals have been examined by the methods outlined above. Some of the crystals were extremely pure, others were doped; the growth directions were either ⟨111⟩ or ⟨100⟩. It was found that crystals grown at the usual growth rates of a few millimetres per minute always contained microdefects, usually in a striated distribution.

We have seen that preferential etching of crystal cross-sections revealed striae of shallow etch pits (fig. 2). X-ray transmission topographs generally showed no contrast. As stated earlier, this suggests that the etch pits are related to very small crystal defects — vacancy clusters or impurity precipitates. Another indication of this was that new pits are continuously formed during etching as the surface recedes. The occurrence of precipitates was very improbable because of the low concentration of impurities in this floating-zone material. The impression that the microdefects in our crystals consisted of vacancy clusters was strengthened by the observation that crystals whose dislocation density was above a particular value (about 1000 cm^{-2}) did not



a

b



c

Fig. 7. Topographs of lithium-decorated crystals. *a*) Detail of a region at the edge of a silicon crystal. Magnification 13×. *b*) Fragment of a germanium crystal with some striations of vacancy clusters. Magnification 50×. *c*) Fragment of a germanium crystal with a more homogeneous distribution of vacancy clusters. Magnification 50×.

exhibit microdefects, either on etching or on topographs. This was true for both silicon and germanium.

In silicon two types of microdefect were found which differ in size and concentration. The average concentrations of the larger type (called A-clusters) were between 5×10^5 and 10^7 cm^{-3}; they were determined by counting their images in the X-ray topographs of decorated crystals. Since it was not found to be possible to decorate the smaller clusters, the B-clusters, with lithium [10], nor with copper in the crystal regions where the concentration of the A-clusters was greater than 10^5 cm^{-3}, we attempted to estimate the concentrations of the smaller defects from the ratio between the numbers of smaller and larger etch pits; see *fig. 8*. Their concentration was found to be between 10^7-10^8 cm^{-3}.

Fig. 8. Preferentially etched silicon with etch pits of various sizes. The etch pits have been made more clearly visible by using interference contrast (Nomarski's method, magnification 150×).

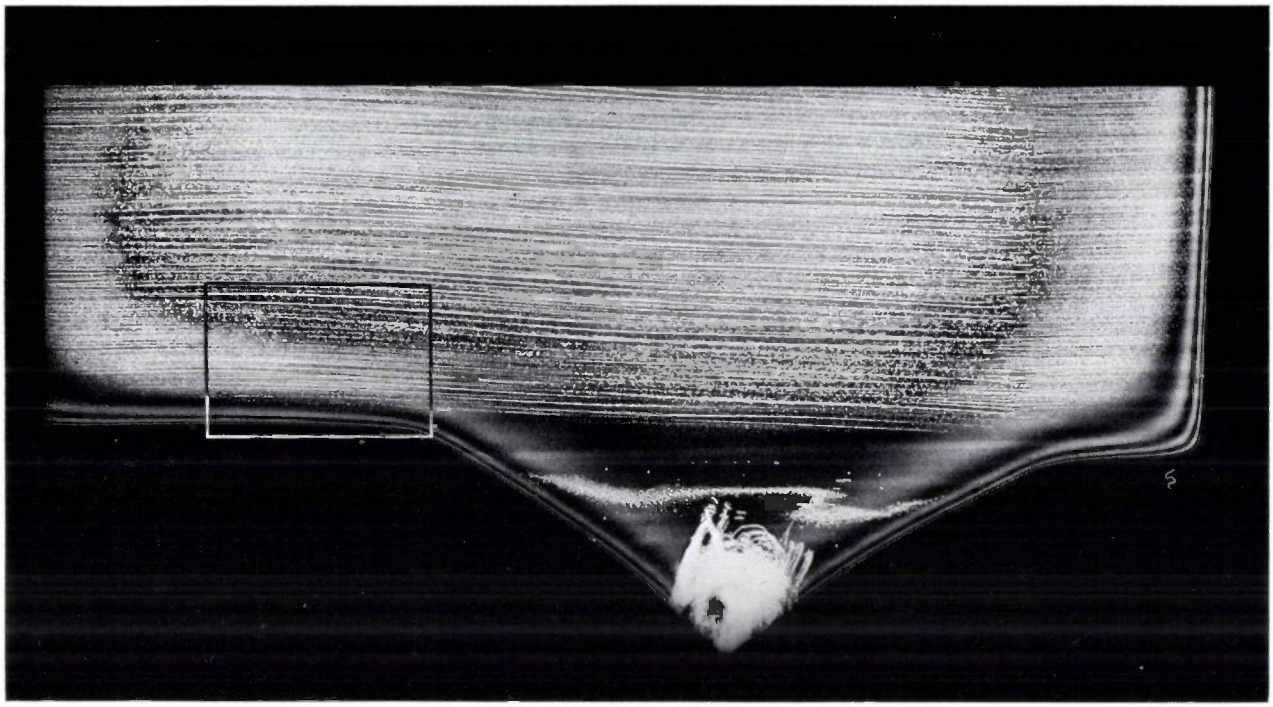
The surface region of crystals grown in argon was only found to contain B-clusters to a depth of about 2 mm. The surface layer of crystals grown in vacuum was completely free of defects.

In dislocation-free *germanium* grown under hydrogen at atmospheric pressure, X-ray topographs of Li-decorated slices also showed microdefects. As with silicon there was a surface layer of a few millimetres with a low concentration of defects. In the bulk of the crystals the defects are often distributed as striations but sometimes a more random distribution is found. The concentration of defects was between 3×10^6 and 5×10^7 cm^{-3}.

In order to get information about the defect distribution present in the crystal *during* growth, Si crystals were quenched after a certain growth period. This was done by a rapid separation of crystal and melt so that the crystal cooled from the melting point (1420 °C) to below 700 °C within 50 s. Analysis of such quenched crystals (see *fig. 9*) showed that the B-clusters are formed first with subsequent formation of the larger A-clusters. The temperature at which the cluster formation starts depends on the growth rate (see *Table II*).

It remains to discuss the mechanism of nucleation of the vacancy clusters and to explain the effect just mentioned in which the small clusters are formed first and the large ones later.

[10] A. J. R. de Kock and P. G. T. Boonen, J. appl. Phys. **44**, 2816, 1973.



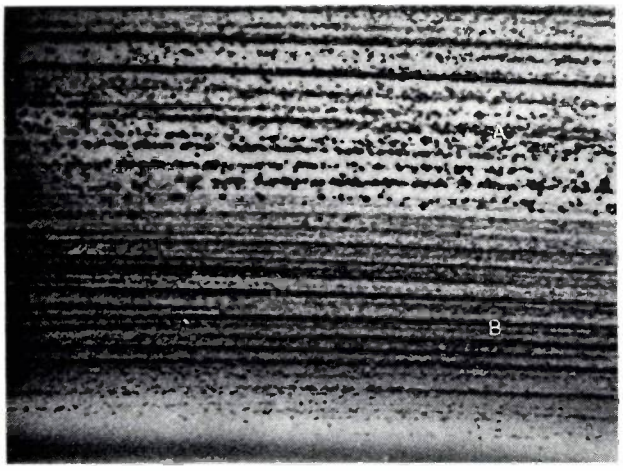


Fig. 9. Crystal quenched from the melt to obtain information about the distribution of defects during growth. *Left:* an enlarged picture of a region near the lower edge. At *A* most of the defects are large vacancy clusters (type A). At *B* only the smaller clusters of type B are found.

Nucleation of the vacancy clusters

The diffusion coefficient of vacancies in silicon is high and of the order of 10^{-5} cm^2 s^{-1} [4]. However, the vacancy clusters are not distributed evenly throughout the crystal, but in a sharply defined striated pattern. This suggests that vacancy clusters are not formed via homogeneous nucleation but that some slow-diffusing and inhomogeneously distributed impurity plays a part in the nucleation process. It is known that stable complexes of vacancies with oxygen can arise in a silicon crystal [11]. We now assume [12] that certain types of such complexes can function as nuclei for the condensation of vacancies in a heterogeneous nucleation process. This give the following picture. During cooling of the crystal, vacancy-oxygen complexes are formed, some of which become so large that they start to act as nuclei for vacancy condensation. This process starts when the temperature has fallen below a certain value [4], and results in the formation of the B-clusters. These defects are probably present in the form of voids, but is also possible that they exhibit a superlattice structure in which the vacancies are regularly arranged [13]. On further cooling the B-clusters grow in size because of continuous vacancy condensation. At a second critical temperature the largest clusters become unstable and subsequently change their shape to that of the A-clusters. Most probably this occurs via a collapse process such as was first described by F. Seitz [14], resulting in the formation of stacking faults and dislocation loops. The effect in which the A-clusters can be decorated with lithium, but the B-clusters cannot, again indicates that the A-

Table II. The temperature at which the precipitation of A- or B-clusters starts in a dislocation-free silicon crystal depends on the rate of growth.

Growth rate (mm/min)	T_B (°C)	T_A (°C)
1	1407	1370
3	1350	1050

and B-clusters differ not only in size but indeed also in character.

The crystal surface region will exhibit a less marked tendency for the formation of vacancy-oxygen complexes for a variety of reasons: the surface acts as a vacancy sink and the oxygen concentration will be lower near the surface because of SiO evaporation at the surface of the melt. The cooling rate near the surface is also greater than in the bulk of the crystal, so that there is less time for the formation of complexes there. In crystals grown in argon these processes cause complete elimination of the A-clusters at the surface (see figs. 2 and 6), while B-clusters are still formed. The surface layer of crystals grown in vacuum, however, is completely free of microdefects (*fig. 10*). This is probably related to the higher evaporation rate of SiO in vacuum.

The clusters in germanium crystals are also usually distributed in a striated pattern, indicating that the existence of a similar nucleation process to that in silicon can be assumed. Sometimes the defects in germanium are distributed more uniformly (fig. 7*c*). This is not really surprising, since the temperature gradients in the Czochralski process are smaller than those in the floating-zone process.

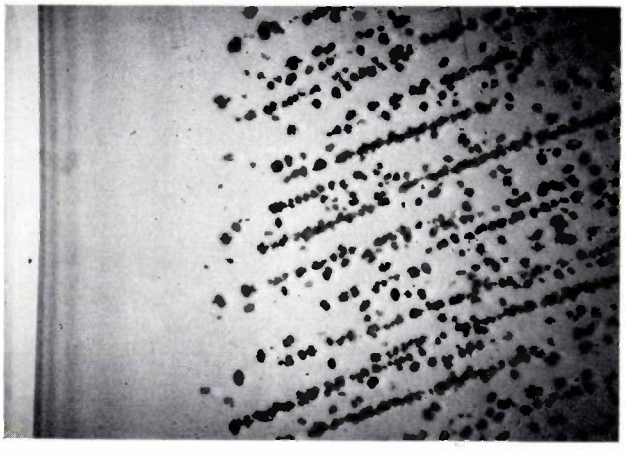


Fig. 10. X-ray topograph of a silicon crystal grown in vacuum and decorated with copper. There are no B-clusters at the edge. Magnification 20×.

It should be noted that the pattern of striations formed by the clusters in silicon is of far more significance than the small periodic variations in the oxygen concentration due to variations in the thermal symmetry. The concentration pattern of the clusters may even contain sharp equidistant peaks, separated by regions in which the concentration is almost zero. This happens because the nucleation starts at places where the oxygen concentration is greatest. These nuclei 'draw' the very mobile clusters to them, and in this way deplete their surroundings before other nuclei are able to form there.

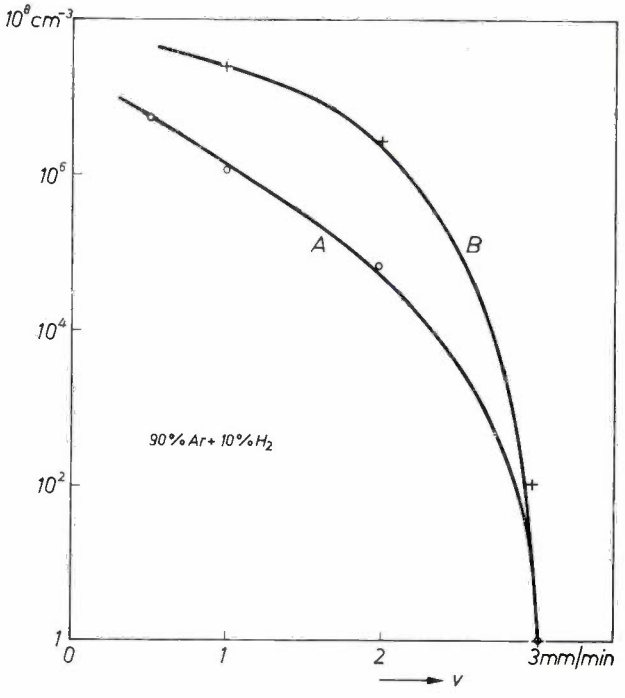


Fig. 11. Concentration of A- and B-clusters as a function of the growth rate for silicon crystals, grown in an atmosphere of argon with 10% hydrogen.

Elimination of vacancy clusters

As supersaturation of vacancies in the growing crystal cannot be avoided, the only possibility of eliminating clusters is to prevent their nucleation. This can only be done by inhibiting the formation of the nuclei — the vacancy-oxygen complexes. Three methods of doing this will now be described: reduction of the oxygen concentration, hydrogen doping and rapid growth.

Reduction of oxygen concentration

Clearly there will be a lower probability of formation of vacancy-oxygen complexes if the oxygen concentration can be reduced. To achieve this extremely pure starting material has to be used. To avoid contamination very clean growth conditions are required, i.e. the crystal must be grown by the floating-zone method in ultra-high vacuum.

In addition, reduction in the periodic variation of the oxygen concentration also reduces the probability of formation of complexes. The growth rate should therefore be kept as constant as possible. This means that particular attention should be paid to the thermal rotational symmetry during growth.

[11] J. W. Corbett, G. D. Watkins and R. S. McDonald, Phys. Rev. **135**, A 1381, 1964.
[12] A. J. R. de Kock, Appl. Phys. Letters **16**, 100, 1970, and J. Electrochem. Soc. **118**, 1851, 1971.
[13] Private communication from Drs J. Hornstra of these Laboratories.
[14] F. Seitz, Phys. Rev. **79**, 890, 1950.

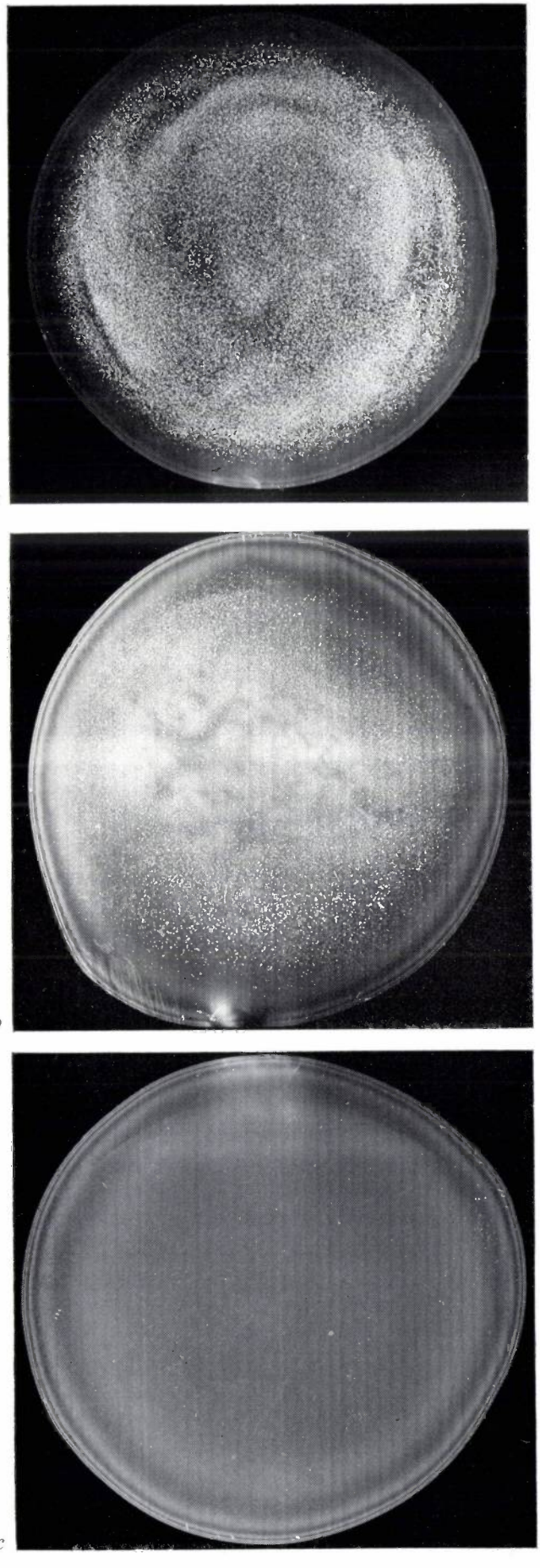


Fig. 12. X-ray topographs of slices from a silicon crystal, grown in an atmosphere of argon with 10% hydrogen. Diameter about 25 mm. *a*) Growth rate 1 mm/min. Concentration of the A-clusters 8×10^5 cm^{-3}. *b*) Growth rate 2 mm/min. Concentration of the A-clusters 8×10^4 cm^{-3} and of the B-clusters 10^6 cm^{-3}. *c*) Growth rate 3 mm/min. At this growth rate no clusters were formed.

Although these measures do indeed reduce the formation of vacancy clusters, they are not entirely eliminated.

Hydrogen doping

Another way of preventing the formation of the vacancy-oxygen complexes is to tempt the oxygen component away to some more attractive partner, introduced as another impurity. Such a liaison reduces the mobility of the oxygen and thus reduces the probability of formation of vacancy-oxygen complexes.

To be effective this impurity should meet the following requirements:

a) It must be possible to introduce it into the crystal in sufficient quantity.

b) Its affinity for oxygen in the silicon lattice should be at least comparable with the affinity of vacancies for oxygen.

c) Because the oxygen-impurity reaction should take place before vacancy-oxygen association starts, the impurity diffusion coefficient should be greater than that of oxygen.

d) The electrical properties of the crystal should not be affected by the doping; the impurity should not therefore have the character of a donor or acceptor.

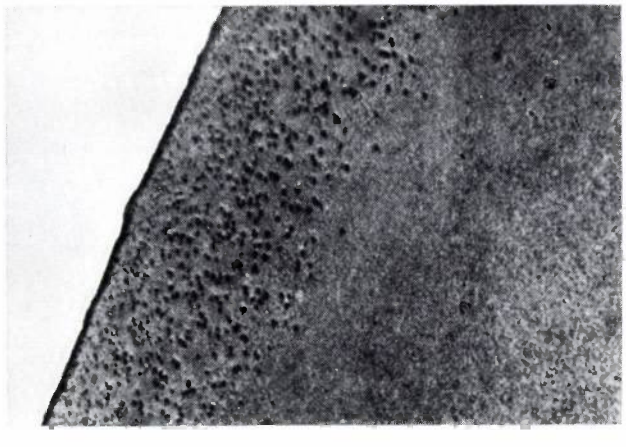
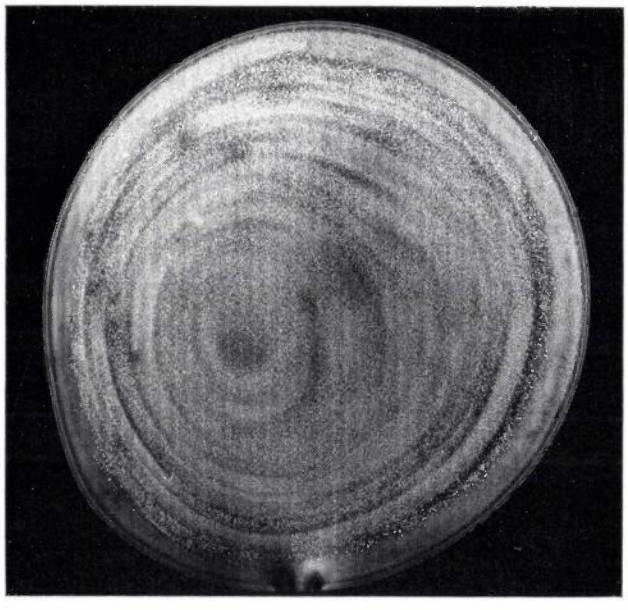
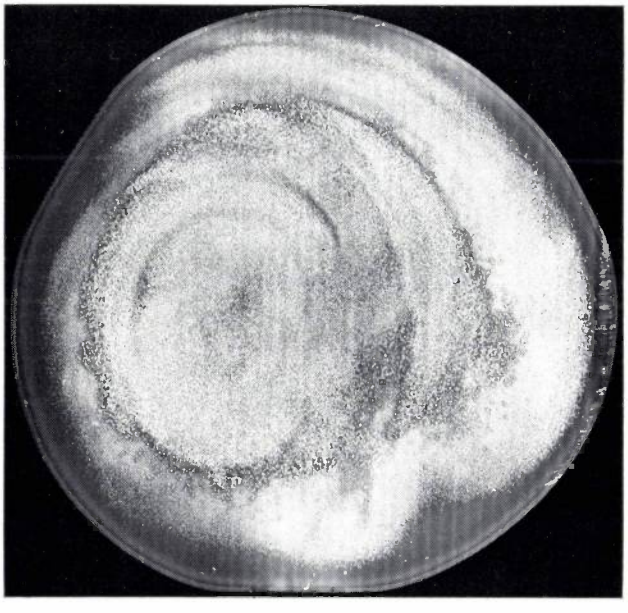


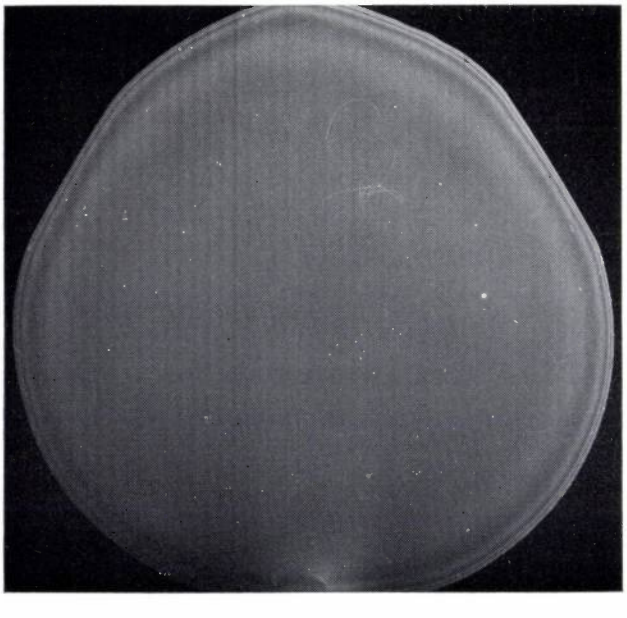
Fig. 13. Lithium precipitates in the edge region of a dislocation-free germanium crystal grown in a hydrogen atmosphere and free of dislocations and vacancy clusters. Magnification 70×. The high concentration at the edge of the crystal suggests precipitation of hydrogen.



a



b



c

Fig. 14. X-ray topographs of copper-decorated slices of silicon crystals grown at various rates in an atmosphere of pure argon. *a*) Growth rate 3 mm/min. Concentration of A-clusters 1.1×10^{6} cm^{-3}, concentration of B-clusters 3×10^{7} cm^{-3}. *b*) Growth rate 4 mm/min. Concentration of A-clusters 1.3×10^{5} cm^{-3}, concentration of B-clusters 10^{6} cm^{-3}. *c*) Growth rate 5 mm/min. No clusters formed.

The only impurity that meets all these requirements is hydrogen. Accordingly, a number of silicon crystals were grown in argon with an addition of 10% hydrogen, and a study was made of the way in which the cluster formation was related to the growth rate of the crystal. The result is shown in *fig. 11*. At a growth rate of only 3 mm/min completely cluster-free material (*fig. 12*) is obtained.

Such cluster-free crystals, however, are not entirely free of defects. Low-temperature lithium decoration revealed the presence of a new type of microdefect, which was found to dissolve at temperatures above 500 °C [4]. This meant that they could not be detected by copper decoration. These microdefects are hydrogen precipitates formed as a result of supersaturation during cooling of the crystal. Precipitate concentrations up to 10^{11} cm^{-3} have been detected. At such concentrations the material becomes brittle. It was found that the precipitate concentration exhibits a maximum near the crystal surface.

A similar thin surface region containing a high concentration of lithium precipitates has been found in germanium crystals grown in hydrogen and then decorated with lithium, as can be seen in *fig. 13*. This suggests that hydrogen precipitation also takes place in germanium.

Method of rapid growth

We now come to the third method for the elimination of vacancy-cluster formation. The rate at which oxygen-vacancy complexes are formed will be diffusion-limited When the cooling rate of the crystal is sufficiently increased, less time is available for the formation of these complexes and fewer nuclei for the precipitation of vacancies are therefore formed. The cooling rate of a crystal can be speeded up by increasing the growth rate or by increasing the temperature gradients in the crystal during growth. The temperature gradients in the growing crystal can be increased by using a gas atmosphere with a large thermal conductivity (for instance He) and a high pressure or by introducing forced convection along the crystal.

The effect of a high rate of growth was predicted from observations on crystals grown at normal rates

— up to 3 mm/min — in argon, with cooling rates of up to 1.5 °C s^{-1}. In the normal regions of these crystals there are clusters in the normal concentrations (about 10^6 cm^{-3}). The dislocation-free region of the thin neck at the top end of the crystal, however, grown at a high rate to eliminate line defects, does not contain vacancy clusters. Diffusion calculations showed that, in spite of the small diameter of this region, the number of vacancies capable of reaching the surface during cooling of the neck is negligible, because of the fast cooling rate of this region (25 °C s^{-1}). This indicated that cluster formation could be suppressed by raising the cooling rate to a value between 1.5 and 25 °C s^{-1}. It has indeed been found that the cluster concentrations could be reduced to zero by increasing the growth rate to values of 5 mm/min or larger (*fig. 14* and *15*).

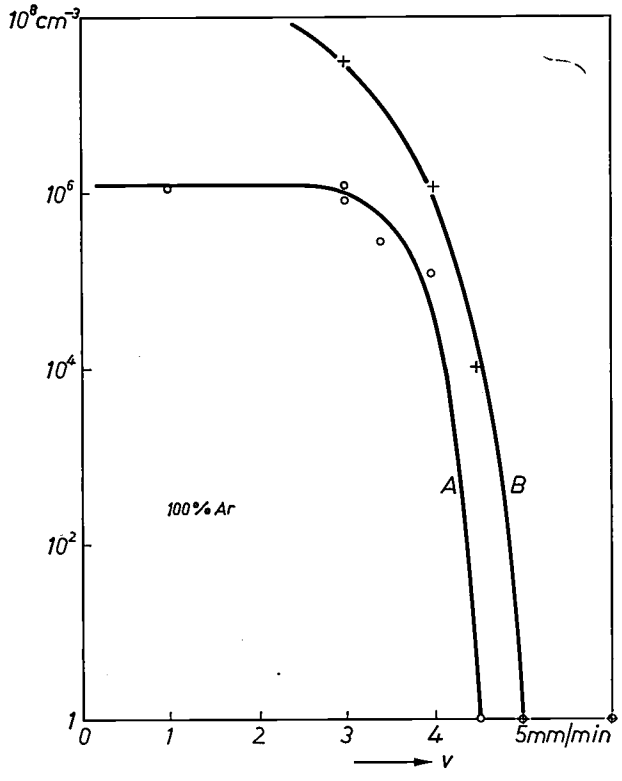


Fig. 15. Concentrations of A- and B-clusters as a function of the growth rate for silicon crystals grown in an atmosphere of pure argon.

Of the three methods for preventing the formation of vacancy clusters, raising the growth rate has in general proved the most satisfactory. The performance of semiconductor devices that are sensitive to the presence of vacancy clusters is greatly improved if they are made of cluster-free material. For example, the white spots that spoil the picture given by a silicon vidicon (fig. 1) are completely eliminated.

Summary. High-purity dislocation-free crystals of Si and Ge can be grown from the melt (Si by the floating-zone method; Ge usually by the Czochralski technique). Nevertheless these crystals are not absolutely perfect, since they contain vacancy clusters, formed as a result of supersaturation and precipitation of vacancies. The presence of these microdefects in silicon can adversely affect the electrical performance of some silicon devices. The vacancy precipitation occurs at nuclei consisting of stable vacancy-oxygen complexes. This explains why the vacancy clusters are usually distributed in a striated pattern. The formation of the clusters can be suppressed if the probability of nucleation is reduced by reducing the oxygen concentration, doping with hydrogen or raising the growth rate. Silicon crystals can be grown that are completely free of vacancy clusters.

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: The growth and the degree of dispersion of in situ composites.
Acta Electronica **17**, 75-86, 1974 (No. 1). *E*

V. Belevitch: Lorentz transformations and equivalent networks.
Philips Res. Repts. **29**, 214-242, 1974 (No. 3). *B*

G.-A. Boutry: Colloque sur la photoémission (introduction).
J. Physique **34**, C6/1-4, 1973 (Colloque C6). *L*

J. W. Broer: Do you write well?
IEEE Trans. **PC-16**, 42, 1973 (No. 2). *E*

M. Brouha & K. H. J. Buschow: The pressure dependence of the Curie temperature of rare earth - cobalt compounds.
J. Physics F **3**, 2218-2226, 1973 (No. 12). *E*

V. Chalmeton & G. Eschard: Reduction of the relative variance of the single-electron response at the output of a microchannel plate.
Adv. in Electronics & Electron Phys. **33A**, 167-174, 1972. *L*

G. Clément: An ultra-fast shutter tube for exposure times below 0.5 nanosecond.
Adv. in Electronics & Electron Phys. **33B**, 1131-1136, 1972. *L*

C. D. Corbey & R. A. Gough (University of Bradford): The effects of package and mount parameters on wide-band varactor tuned oscillators.
4th Biennial Cornell Conf. on Microwave semiconductor devices, circuits, and applications, Ithaca, N.Y., 1973, pp. 165-175. *M*

J. P. Deschamps: On a theory of discrete functions, Part III. Decomposition of discrete functions.
Philips Res. Repts. **29**, 193-213, 1974 (No. 3). *B*

H. J. A. van Dijk, J. Goorissen, U. Gross, R. Kersten & J. Pistorius: Crystal diameter control in Czochralski growth.
Acta Electronica **17**, 45-55, 1974 (No. 1). *E, A*

H. Dötsch, H. J. Schmitt & J. Müller: Detection and generation of magnetic bubble domains using ferrimagnetic resonance.
Appl. Phys. Letters **23**, 639-641, 1973 (No. 11). *H*

G. Eschard & P. Dolizy: Contribution à l'optimisation des photocathodes trialcalines semi-transparentes.
J. Physique **34**, C6/61-63, 1973 (Colloque C6). *L*

A. N. Godard: Design of optimum FIR bandpass filters through frequency transformations.
Philips Res. Repts. **29**, 243-252, 1974 (No. 3). *B*

J. J. Goedbloed: Intrinsic AM noise in singly tuned IMPATT diode oscillators.
IEEE Trans. **ED-20**, 752-754, 1973 (No. 8). *E*

J. Graf, M. Fouassier, R. Polaert & G. Savin: Characteristics and performance of a microchannel image intensifier designed for recording fast luminous events.
Adv. in Electronics & Electron Phys. **33A**, 145-152, 1972. *L*

G. Haas: Über die Druckgeschwindigkeit mechanischer Drucker mit kontinuierlich bewegtem Typenträger.
Feinwerktechnik + Micronic **77**, 354-359, 1973 (No. 8). *H*

J. Hallais: Epitaxie en phase vapeur des composés pseudo-binaires III-V.
Acta Electronica **17**, 19-31, 1974 (No. 1). *L*

N. Hazewindus & F. Martis: On the measurement of the field of a magnetic quadrupole lens by means of a Hall probe.
Nucl. Instr. Meth. **112**, 611-613, 1973 (No. 3). *E*

H. Hervet (Collège de France, Paris), **J. P. Hurault & F. Rondelez:** Static one-dimensional distortions in cholesteric liquid crystals.
Phys. Rev. A **8**, 3055-3064, 1973 (No. 6). *L*

E. P. Honig: 1/*f*-noise of bodies of arbitrary shape and of point contacts.
Philips Res. Repts. **29**, 253-260, 1974 (No. 3). *E*

D. R. Hunter (University of Oxford), **D. H. Paxman, M. Burgess & G. R. Booker** (Univ. Oxford): Use of the SEM for measuring minority carrier lifetimes and diffusion lengths.
Proc. Conf. on Scanning electron microscopy: systems and applications 1973, Newcastle upon Tyne, pp. 208-213. *M*

W. P. A. Joosen: Helical pipe flow in the presence of an axial pressure gradient.
Philips Res. Repts. **29**, 383-400, 1974 (No. 4). *E*

D. Kasperkovitz: A bipolar four-phase dynamic shift register.
IEEE J. **SC-8**, 343-348, 1973 (No. 5). *E*

S. K. Kurtz: Phase transitions between optically distinguishable states and some potential applications.
Conf. on Phase transitions and their applications, University Park, Pa., USA, 1973, pp. 29-59. *N*

D. Meyer-Ebrecht: Digitalisierung und Verarbeitung frequenzanaloger Kraft- und Gewichts-Meßsignale.
VDI-Berichte No. 202, 53-57, 1973. *H*

E. J. Millett, J. A. Morice & J. B. Clegg: The computer evaluation and interpretation of photographically recorded spark source mass spectra.
Int. J. Mass Spectrom. Ion Phys. **13**, 1-24, 1974 (No. 1). *M*

J. Monin (Conservatoire National des Arts et Métiers, Paris) **& G.-A. Boutry:** La photoémission des métaux alcalins purs. Résultats et essais d'interprétation.
J. Physique **34**, C6/13-17, 1973 (Colloque C6). *L*

J. J. Opstelten, D. Radielović & W. L. Wanmaker (Philips Lighting Division, Eindhoven): The choice and evaluation of phosphors for application to lamps with improved color rendition.
J. Electrochem. Soc. **120**, 1400-1408, 1973 (No. 10).

G. den Ouden (Philips Welding Electrode Factory, Utrecht): Microparticles in mild steel weld metal.
Welding J. **51**, 542s-543s, 1972 (Res. Suppl., Nov.).

R. C. Peters & E. André (RTC La Radiotechnique-Compelec, Caen): Morphology and thickness of GaP layers and composition of $Ga_{1-x}Al_xAs$ layers grown by liquid phase epitaxy.
Acta Electronica **17**, 9-17, 1974 (No. 1). *E*

G. Piétri: Les tubes électroniques à cathode photoémissive. Instruments de physique expérimentale.
J. Physique **34**, C6/67-77, 1973 (Colloque C6). *L*

L. G. Pittaway: The application of ion storage in electron space-charge fields to the design of a U.H.V. gauge and mass-spectrometer ion source: Part I. The design of a new extractor gauge for U.H.V. pressure measurements, Part II. The construction and performance of the extractor gauge, Part III. A high-sensitivity ion source for mass-spectrometer applications.
Philips Res. Repts. **29**, 261-282, 283-302, 363-382, 1974 (Nos. 3 & 4). *M*

C. J. M. Rooymans: Growth and application of single crystals of magnetic oxides.
Acta Electronica **17**, 33-44, 1974 (No. 1). *E*

D. J. Schipper, T. W. Lathouwers & C. Z. van Doorn: Thermal decomposition of sodalites.
J. Amer. Ceramic Soc. **56**, 523-525, 1973 (No. 10). *E*

H. Scholz: On crystallization by temperature-gradient reversal.
Acta Electronica **17**, 69-73, 1974 (No. 1). *A*

J. W. Slotboom: Computer-aided two-dimensional analysis of bipolar transistors.
IEEE Trans. **ED-20**, 669-679, 1973 (No. 8). *E*

J. L. Sommerdijk: Influence of the host lattice on the infrared-excited blue luminescence of Yb^{3+},Tm^{3+}-doped compounds.
J. Luminescence **8**, 126-130, 1973 (No. 2). *E*

D. B. Spencer & K. G. Freeman: Television broadcasting from satellites.
Wireless World **79**, 607-610, Dec. 1973, & **80**, 39-44, March 1974. *M*

A. L. N. Stevels & A. D. M. Schrama-de Pauw: Vapour-deposited CsI:Na layers: I. Morphologic and crystallographic properties, II. Screens for application in X-ray imaging devices.
Philips Res. Repts. **29**, 340-352, 353-362, 1974 (No. 4). *E*

A. Thayse: On a theory of discrete functions, Part IV. Discrete functions of functions.
Philips Res. Repts. **29**, 305-329, 1974 (No. 4). *B*

W. Tolksdorf: Magnetic garnet single crystal growth from fluxed melts.
Acta Electronica **17**, 57-67, 1974 (No. 1). *H*

J. D. B. Veldkamp: The elastic moduli of particle-filled materials.
J. Physics D **6**, 2012-2024, 1973 (No. 17). *E*

C. H. F. Velzel: Fourier spectroscopy without a computer.
Philips Res. Repts. **29**, 330-339, 1974 (No. 4). *E*

H. W. Werner: Analysis of thin surface layers by secondary ion mass spectrometry and similar methods. Quantitative analysis with electron microprobes and secondary ion mass spectrometry, Proc. Conf. Jülich 1972, pp. 305-342; 1973. *E*

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 10

The Opthycograph

A. G. Bouwer, R. H. Bruel, H. F. van Heek, F. T. Klostermann and J. J. 't Mannetje

In 1967 work on the development of an exceptionally accurate optical pattern generator was initiated at Philips Research Laboratories by F. T. Klostermann and G. C. M. Schoenaker. A machine of this type was considered to be necessary — and has indeed proved to be so — in more advanced work on integrated circuits and such devices as the index tube for colour television. The result of this development, the 'Opthycograph' described here, is the offspring of a happy combination of traditional lines of research at Philips Research Laboratories. The knowledge and experience gained with devices such as hydrostatic bearings proved to be particularly useful.

The development of the Opthycograph was the work of a team of specialists from various different fields. A. G. Bouwer was responsible for the mechanical design and construction, R. H. Bruel devised the control systems using a small computer and produced the hardware and software, while J. J. 't Mannetje designed the hydraulic drive and the associated servo systems. F. T. Klostermann was the man responsible for the complete project, including the general design of the machine and the optical and photographic activities. H. F. van Heek took over this function when the Opthycograph reached the stage of being made operational.

Introduction

A photomask is a pattern of transparent and opaque areas on a transparent substrate. The masks are primarily used in the production of integrated circuits and for making printed circuits. Other uses include the application of phosphor patterns to the screen of the index tube, a type of colour television picture tube.

Formerly a photomask for an integrated circuit was made by starting with a master, possibly drawn by hand, which was then reduced photographically. The master was made large enough for the dimensional accuracy achieved in the drawing to give the required accuracy in the reduced mask.

In masks for integrated circuits the pattern for a single circuit has to be repeated many times, so that one mask can be used for simultaneously forming a large number of identical circuits on a single silicon wafer. For this repetitive process a step-and-repeat camera was developed in our laboratories [1].

The 'cut-and-strip' technique was later introduced for making the masters. In this technique a plastic sheet is used consisting of two layers, the lower one transparent and the upper one opaque. The outside edges of the areas that must be transparent in the original are cut in the upper layer — e.g. with a numerically controlled machine — and the opaque material is then stripped away from these areas.

With the trend towards larger and more complex integrated circuits, we are faced with a considerable increase in the number of elements per circuit, so that larger and more complicated originals have to be made. What is more, the individual elements are tending to become smaller, which makes even greater accuracy necessary. The cut-and-strip technique, which cannot be done automatically, thus presents more difficulties and checking of the work becomes an intractable problem.

A way out of this difficulty is to create the masters entirely automatically. This means that filled-in areas

A. G. Bouwer, R. H. Bruel, Ir F. T. Klostermann and Ir J. J. 't Mannetje are with Philips Research Laboratories, Eindhoven. Drs H. F. van Heek, formerly with Philips Research Laboratories, is now with Philips Elcoma Division, Eindhoven.

[1] F. T. Klostermann, Philips tech. Rev. **30**, 57, 1969.

Fig. 1. The Opthycograph, with the control desk on the right and the actual pattern-drawing part on the left. The slide with the photographic plate can be seen on the left-hand side of the drawing machine and behind it the linear hydraulic motor that drives the slide. Situated on the right of this slide is the slide with the optical column, containing separate projection systems for line drawing and for the flashing-in of frequently recurring details. The drive for this slide is on the far right. The carriage slideways are located under the dark cover bearing the name of the machine. Telescopically sliding parts in this cover ensure that the slideways, which are entirely covered with oil from the slide bearings, always remain completely enclosed. During the drawing operations the photographic plate lies on its carriage unscreened, and the machine is therefore operated in a completely dark and dust-free room.

will have to be drawn and not just the outside edges. As can be seen from *Table I*, there are various methods of doing this, and some of them have already been adopted in various parts of the world [2].

At Philips Research Laboratories we have designed a machine that draws photomasks optically by means of a continuously moving beam of light on a photographic plate. The plate and the optical system that generates the beam move on slides with on hydrostatic bearings, along two slideways at right angles. The slides are driven by linear hydraulic motors, and the movements are controlled by commands from a small computer. The machine, shown in *fig. 1*, has been given the name 'Opthycograph', from *opt*ical *hy*draulic *co*mputer-controlled *graph*ic machine.

The Opthycograph was primarily designed for drawing integrated-circuit masks required for the step-and-repeat camera mentioned above. *Fig. 2* shows an example of such a mask drawn by the Opthycograph. The machine will also produce excellent curved lines (such as the pattern on p. 269). However, the computer control gives the machine such flexibility that it can also be used to make accurate drawings for many other purposes. The flexibility is enhanced by the use of a readable code for short control programs that can be fed in on punched tape. This code allows programs to be verified quickly and makes it easier to apply minor modifications. For the very large control programs, however, the Opthycograph uses magnetic tape as well as punched tape for data input. The magnetic tape always originates from a large computer used for controlling the layout of the elements of integrated circuits. Such data is not directly readable, and can only be modified using the large computer.

Table I. Survey of the various methods of making photomasks. The methods used in the Opthycograph have been underlined.

- Mechanical method
- Photographic plate
  - Electron beam
  - Light beam
    - Stationary lens
    - Moving lens
      - Flash source
      - Continuous light

We shall first of all describe the process of drawing with a light beam and deal with the mechanical construction of the Opthycograph, indicating some of the considerations that played a part in the design of the machine. We shall then describe the control system that ensures that the light beam follows the desired path with the required high accuracy, and finally we shall discuss some details of the computer control.

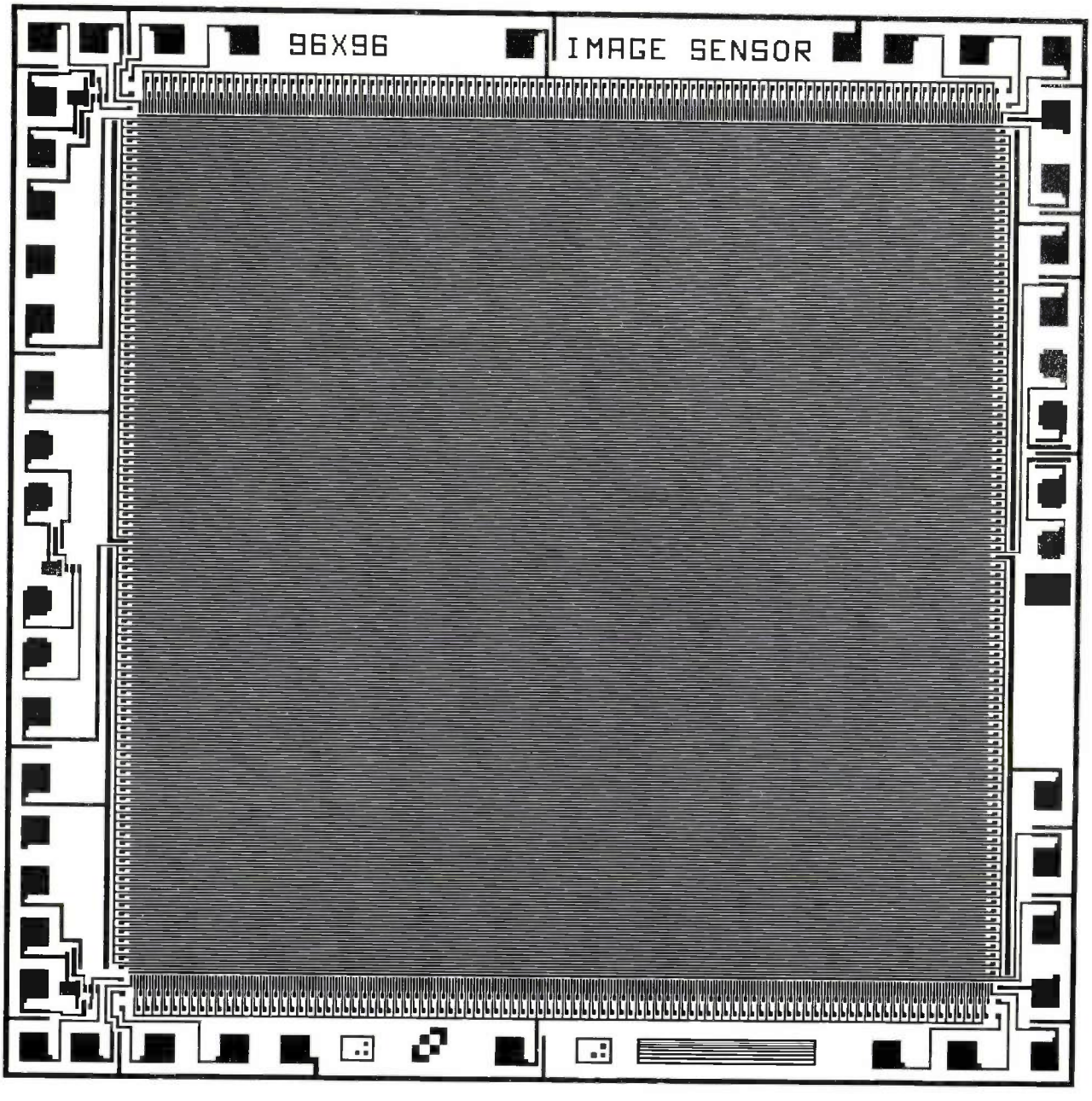


Fig. 2. Example of a mask drawn by the Opthycograph, giving the pattern of conductors for a two-dimensional shift register with 96 × 96 elements. A shift register of this type is used as an image sensor [3]. The mask shown here has been drawn as a pattern of 36 × 36 mm. The long paths were 75 μm wide with a gap of 25 μm between them. In reproduction by the step-and-repeat camera all the dimensions are reduced ten times. It took 51 minutes to draw the complete mask. A photographic enlargement of the mask has been used for this illustration.

Drawing with a light beam

In drawing the transparent and opaque areas that form a photomask, the position of the boundaries between these areas must first of all be accurately defined. Apart from the optical quality of the lenses that form the beam, the shape of the light spot produced by the beam on the plate and the nature of the photographic material used are all important here.

The photographic material we use is a Lippmann plate. This has a grain so fine that more than a thousand line pairs per millimetre can be drawn on it. To obtain a very steep light-to-dark transition along the edge of a line, and thus define accurately the position of the edge, we use a rectangular light spot (*fig. 3*). A dis-

[2] Particulars on this subject can be found in the November number of Bell Syst. tech. J. **49**, 1970, in Proc. Topical meeting on the use of optics in microelectronics, Las Vegas 1971, and in E. Kutzer and R. Martin, Z. ind. Fertig. **62**, 71, 1972.

[3] This application of a shift register is mentioned in F. L. J. Sangster: The 'bucket-brigade delay line', a shift register for analogue signals, Philips tech. Rev. **31**, 97-110, 1970.

advantage of a rectangular spot as compared with a round one is that the orientation of the rectangle must be adapted to the direction of the line.

Another point requiring attention is the exposure. At every point of a line or surface it is necessary to have the same exposure E, irrespective of the speed at which the light spot moves. The exposure is determined by the luminous intensity I, the speed v of the light spot and the length l of the spot. If v is constant, $E = Il/v$. It would be possible to keep E constant at any speed by varying the intensity. However, when the spot was accelerated or slowed down this would cause a variation in E over a distance l at the beginning and end of the line. Limiting this distance by choosing a small l would entail either a high luminous intensity or a low speed, neither of which is desirable. It is however possible to obtain a good result by varying the length of the spot while keeping l constant, provided it is done in the right way. We have adopted the following solution for this problem. We control the leading edge and trailing edge of the rectangle independently of one another, and make the movement of the trailing edge follow that of the leading edge with a delay equal to the required exposure time (*fig. 4*).

The optical system

A moving light beam for drawing a photomask on a photosensitive plate can be produced in two ways. The beam-forming optical system and the plate can be made to move in relation to one another on slides travelling along slideways, or the light beam can be moved by means of a tilting mirror in the image field of a stationary lens (*fig. 5*). The second method, even using the most advanced design of lens, inevitably demands a compromise between image sharpness and the size of the available image field.

The moving-lens solution that we have chosen for the Opthycograph has the advantage of avoiding this compromise. It enables us to image small details with great sharpness on a large field. This means, however, that the relative movements of the plate and the optical system must be carried out extremely accurately, while at the same time the speeds and accelerations must be high to permit complex patterns to be produced in an acceptably short time. This combination of requirements sets very tight specifications for the mechanical design and the servomechanisms that control the movements.

In the Opthycograph the photographic plate and the optical system are mounted on two independent slides. The slides are each driven at their centre of mass, which would be impossible using a cross-slide system with either the plate or the optical system mounted on it. This form of drive is important because no inertial

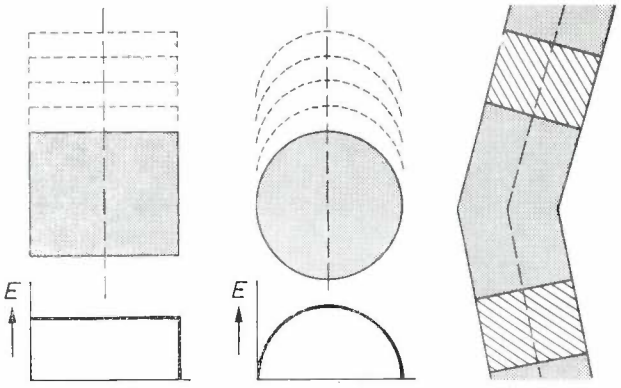


Fig. 3. Distribution of the exposure E over the width of a line; *centre:* drawn with a circular light spot; *left*: drawn with a rectangular light spot. The rectangular spot gives the desired very sharp transition in the exposure at the edge of the line. *Right:* illustrating the need to adapt the orientation of a rectangular light spot to the direction of drawing.

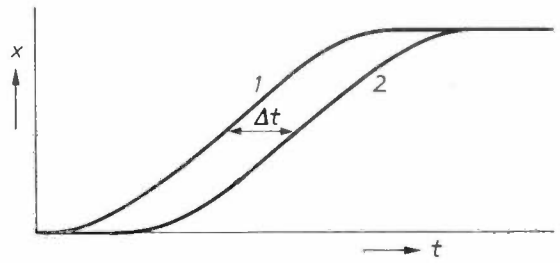


Fig. 4. Drawing with a light spot of variable length. Curves *1* and *2* give the position x of the leading and trailing edges of the rectangle, respectively, as a function of time t. The two edges pass each point on the photographic plate with the same time interval Δt, so that the exposure time is constant.

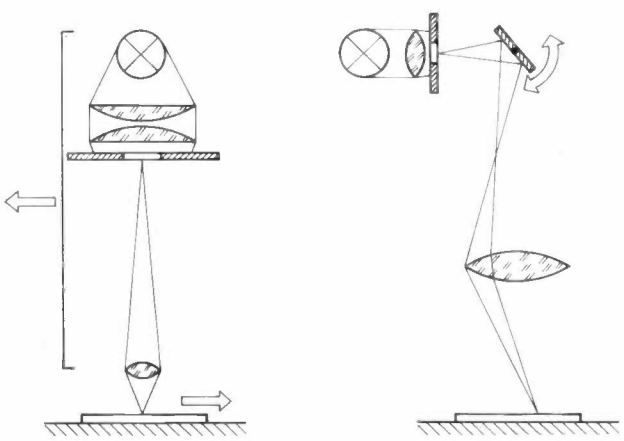


Fig. 5. Drawing with an optical system that moves in relation to the photographic plate (*left*), and with a light beam that moves in the image field of a fixed lens (*right*).

torques arise when the carriages are accelerated or decelerated. The optical system is mounted on the carriage with its axis horizontal and the beam is directed vertically downwards on to the plate by reflection from a mirror. This gives a very compact and rigid assembly with a low-lying centre of mass, giving a minimum of unwanted mechanical vibrations.

The optical system is illustrated schematically in *fig. 6*. A lamp L with a condenser lens illuminates the

slit *S* that shapes the light beam. The slit can be varied both in length and width. The length of the slit can be varied rapidly and continuously to control the exposure, altering the spacing between the two knife-edges *A* and *B*. The movement of *A* is controlled by a solenoid and defines the leading edge of the spot ('light on'). The trailing edge ('light off') is defined by *B*, which is rigidly fixed to the projection column. By appropriately controlling the movable knife-edge and one or both of the slides the exposure control illustrated in fig. 4 is obtained. The width of the slit can be varied in steps, and determines the width of the line produced. Next comes a Péchan prism *P* (*fig. 7*), which can be rotated to match the orientation of the image of the slit to the direction of drawing.

A beam combiner *C* is used in conjunction with a flash source *F* and a photomask *M* for flashing frequently recurring complicated details on to the plate as a single set. Finally, the light beam is directed downwards by a mirror to pass through the actual projection lens *Obj* to the plate *Pl*.

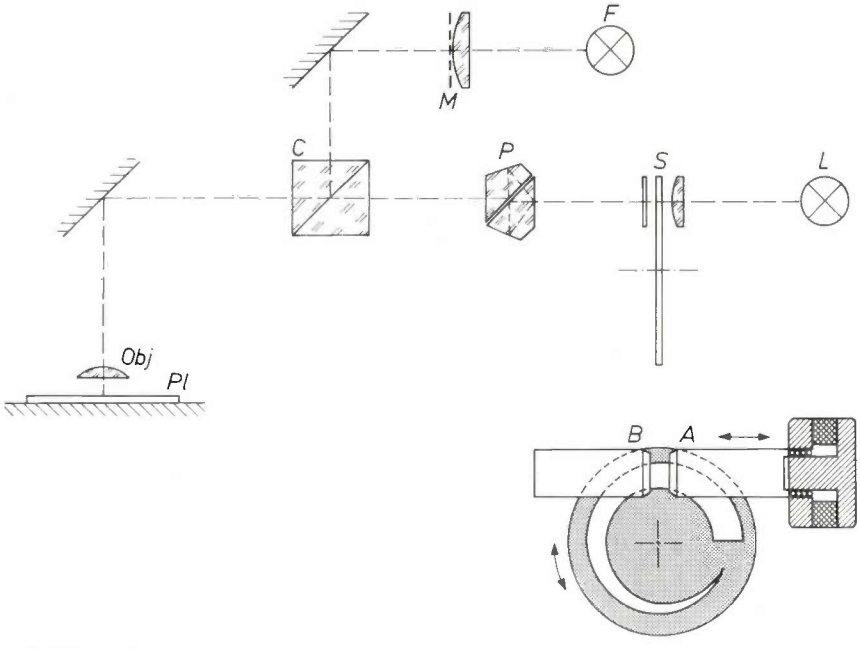


Fig. 6. The optical system of the Opthycograph (schematic). *L* lamp. *S* slit, controlled both in width and length. The slit width is controlled by means of a rotary disc with a slot of variable width. The slit length is determined by the distance between the movable knife-edge *A* and a fixed knife-edge *B*. *P* Péchan prism for rotating the slit image. *C* beam combiner through which frequently recurring details can be projected as a whole on to the plate with the aid of a separate mask *M*. *F* flash source for such details. *Obj* objective. *Pl* photographic plate.

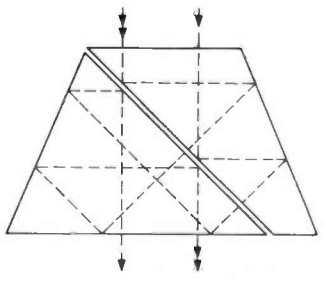


Fig. 7. The Péchan prism. The light rays are reflected an odd number of times, so that the prism is the equivalent of a single mirror; rotation of the prism about a vertical axis through an angle α results in a rotation of the slit image through an angle 2α.

Mechanical construction

The slides are fitted with hydrostatic bearings, which were developed at Philips Research Laboratories [4]. The slideways are made of steel and are secured to a baseplate made of granite, a material of such great geological age that there is no possibility of mechanical instability. The slideways consist of three bars fastened to the base-plates with retaining bolts. Careful adjustment of the bars gives very accurate linear alignment of the slideways, as illustrated in *fig. 8*.

Since there is a continuous flow of oil between the bearing surfaces, there is no metallic contact between them and therefore no dry or Coulomb friction. The only viscous friction present is proportional to the speed of travel of the carriage. Consequently there are no stick-slip effects between the moving parts. The lateral stiffness of this bearing is extremely high.

The slides are driven by linear hydraulic motors on hydrostatic bearings (*fig. 9*). These motors consist of cylinders in which a piston is displaced by oil pres-

[4] H. J. J. Kraakman and J. G. C. de Gast, Philips tech. Rev. **30**, 117, 1969.

sure. They drive the slides directly via the coupling illustrated in *fig. 10*. The operation of the coupling is based on the same principle as that of the hydrostatic bearing. In principle the coupling only transmits forces acting in the axial direction. The axial stiffness is high (2000 N/μm), which is necessary because the coupling is contained in the control loop for the slide drive. Torques or transverse forces due to inaccuracies in the alignment of the slide with respect to the motor do not occur with this type of coupling.

Measurement of the slide position

It is necessary to measure the position of a slide extremely accurately. This is done by an optical measuring system, using a phase grating [6], fitted to the Opthycograph. Analog interpolation can be used to give a measurement accuracy of within 0.1 μm. The effect of differences in expansion between the photographic plate and the gratings is reduced by using the same type of glass as the substrate for the gratings and the photographic emulsion. Furthermore, oil from the slide bearings circulates around the gratings and is kept at constant temperature by a thermostat. The same oil is used for the couplings between motors and slides, so that the motors, which become heated irregularly, are thermally insulated from the slides.

Performance of the Opthycograph

The performance of the Opthycograph, details of which are summarized in *Table II*, reflects the requirements of future users as well as the technological possibilities and constraints at the beginning of the development.

Table II. Performance of the Opthycograph.

Largest photomask	200 × 200 mm
Smallest displacement (step)	0.5 μm
Absolute accuracy	± 2 μm
Repetition accuracy	± 0.5 μm
Maximum drawing speed	10 mm/s
Maximum acceleration	0.5 m/s²
Line-width	2-1900 μm (in 200 steps)
Maximum flashed detail	1500 μm

The minimum line width, the step size and the various accuracies were specified for drawing masks for integrated circuits. Since a light-optical system is used both in the Opthycograph and in the step-and-repeat camera it would be pointless to have lower values than those given here.

Another objective was to make the machine fast enough to produce a series of six photomasks for one IC in one working day. This operating speed was in fact achieved for a circuit consisting of some 200 elements,

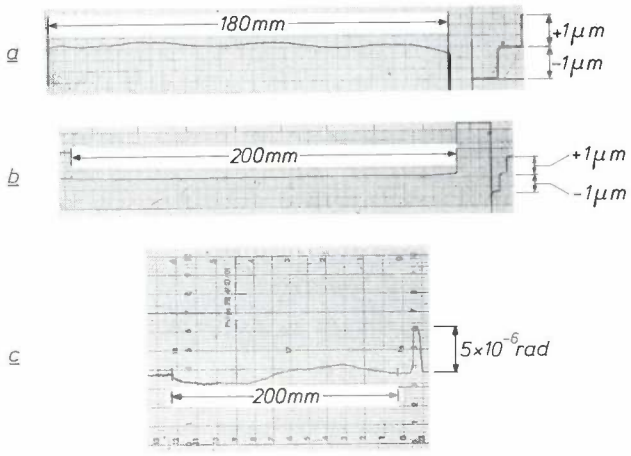


Fig. 8. Accuracy of a slideway. *a*) Lateral deviation of the optical column over a distance of 180 mm; the deviation is measured at the location of the light beam incident on the plate (the working point). *b*) Lateral deviation of the slide with the photographic plate. *c*) Rotation of the slide with the photographic plate about a vertical axis.

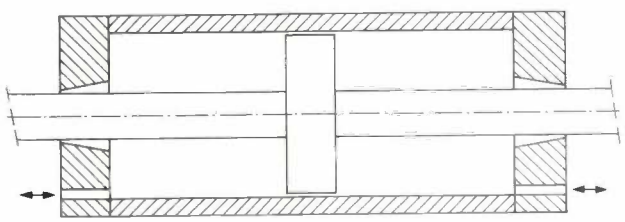


Fig. 9. Cross-section of a linear hydraulic motor. Because the guide apertures are conical the shaft is supported on hydrostatic bearings. The piston has a small clearance in the cylinder. As a result of these measures there is no metal-to-metal contact and therefore no dry friction (Coulomb friction).

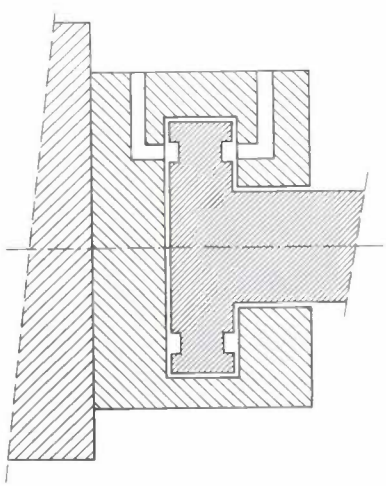


Fig. 10. Hydraulic coupling between linear motor and slide. The shaft driven by the motor ends in a disc with a circular channel on each face. The disc is enclosed with minimum play in a housing fastened to the slide. The two channels are connected to a 40 atm oil-pressure line via a membrane double restrictor [5], an element with internal feedback, giving high axial stiffness. Any non-axial motion of the drive only causes oil circulation in the channels. This circulation meets no opposition, so that the stiffness for such motion is very low.

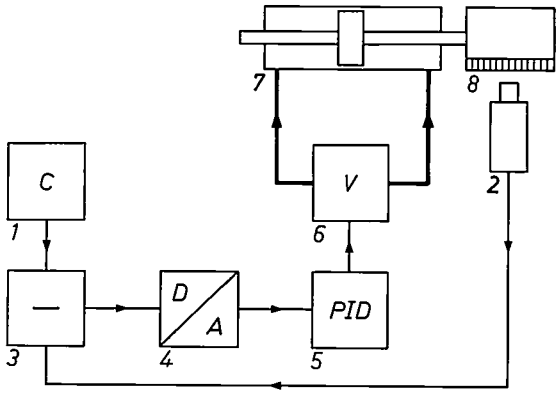


Fig. 11. Basic diagram of the control loop for the displacement of the slides. *1* computer that gives the control commands. *2* system for measuring the slides position. *3* digital subtraction circuit. *4* digital-to-analog converter. *5* automatic controller. *6* servo valve. *7* linear motor. *8* slides.

which at the time was considered to be a normal size. Present-day ICs are many times larger and photomasks for these require correspondingly longer operating times. For masks largely consisting of a repetition of identical details, such as those for memory circuits, the operating time can be reduced considerably by flashing these details in.

Slide speed and acceleration

The maximum speed of a slide is important only when it is necessary to draw a pattern consisting of long lines, as in the case of masks for index-tube screens. The time during which a slide is accelerated is then short compared with the time during which it travels at its maximum speed. In the drawing of IC masks, however, consisting of large numbers of short line segments, the maximum speed is seldom if ever reached. In this case the maximum *acceleration* determines the average speed, and hence the speed of operation.

The maximum slide acceleration that can be achieved depends on the slowest element of the circuit controlling the movement of the slide. In the Opthycograph this is the servo valve in the oil-feed line to the hydraulic motor. When we started to develop the Opthycograph we had no experience in designing and building a valve of the kind required for our purpose, and we therefore used a commercially available model. The characteristics of this valve largely determined the design of the control loop and hence the dynamic characteristics of the whole machine.

The servo valve has this marked influence because the resonant frequencies of the various components must be sufficiently far apart to obtain satisfactory behaviour of the control loop. In designing our motor we therefore ensured that its resonant frequency was higher than that of the valve. Once the dimensions of the motor are fixed, they determine, together with the maximum oil flow rate through the valve, the maximum speed that can be expected. Work is at present being carried out on the design of substantially faster servo valves, and promising results have already been achieved.

The control loops

Two control loops ensure that both slides carry out the required movements with the specified accuracy. In principle the two loops are identical; the only difference is that, because of the different masses of the loads, they have different resonant frequencies: 215 Hz for the slide that carries the photographic plate, and 184 Hz for the slide with the optical system. The behaviour of the control loops depends mainly, however, on that of the servo valves. These are identical in the two loops and have a bandwidth of 150 Hz, so that in practice the difference in resonant frequency has little effect.

The principal components of the control loops (*fig. 11*) are the position-sensing system already mentioned, a controller, the electrohydraulic servo valve and the linear hydraulic motor operated by this valve. The controller has to keep the difference between the desired and the measured position as small as possible. Both positions are given by digital signals, so that the position error is also given by a digital signal. This signal passes through a digital-to-analog converter before reaching the input of the controller.

The controller used is a proportional amplifier with two extra branches in parallel with it, one with an integrating action and the other with a differentiating action (A combination of this type is known as a PID controller.) These additions have the effect of limiting the drift of the motor, while at the same time improving the speed of response to changes of commands and reducing overshoot.

The servo valve (*fig. 12*) has a spool with four orifices, permitting one cylinder chamber to be connected to the high-pressure line and the other one simultaneously to the oil reservoir, or the other way about. The speed of the motor depends on the size of the valve opening, and also on the mechanical load. In the neutral position all four orifices are almost completely closed, so that apart from leakage the oil consumption is virtually zero. As can be seen in the figure, the spool of the valve is operated by the pressure difference obtained by means of an electrically driven flapper-nozzle system. Mechanical feedback is intro-

[5] A description of this membrane double restrictor is given in the section on hydrostatic bearings in the article by Kraakman and De Gast [4].

[6] H. de Lang, E. T. Ferguson and G. C. M. Schoenaker, Philips tech. Rev. 30, 149, 1969.

duced between the flapper and the spool to improve the linearity of the valve.

Various measures had to be taken to make the valve meet our requirements. For example, a supply pressure of 70 atm was chosen, this being the minimum pressure below which the valve can no longer operate properly. Since frictional and hysteresis effects can cause considerable deviations from linearity, especially near the centre position, a 1600 Hz signal is superimposed on the electrical control signal for the valve. This reduces the effect of the friction, and the frequency of 1600 Hz is so high that the motor is unaffected by it.

Feed forward

In addition to the signal representing the slide displacement, the control computer also supplies a signal for the speed at which the line should be drawn (i.e. for the exposure control). This signal is used to improve the response of the control system to a command: as well as a position command the system also receives a signal for the expected speed and acceleration (feed forward). First, however, the speed signal has to be adapted to the characteristics of the control loop. At the same time the speed signal is differentiated to produce the acceleration signal. Finally the position signal from the control computer is delayed a few milliseconds in relation to the speed and acceleration signals to compensate for the inertial effect of the control valve.

The operations to be performed on the speed and acceleration control signals v and a before feeding them to the control loop can be derived from *fig. 13*. In this figure x is the command for the carriage displacement and y the carriage displacement itself and also $\varepsilon = x - y$. The PID controller R has the transfer function H_R, and the transfer function of the mechanical system S is H_S. (The transfer function of a system is the complex frequency-dependent ratio of the output signal to the input signal, i.e. in our case of the carriage position to the current through the valve torque motor.) The signal z is the feed-forward signal, to be derived from the signals for v and a.

We now have: $y = H_S(H_R\varepsilon + z)$. If the control is perfect then the error ε is always equal to zero, and thus $x = y$, so that we then find $x = y = H_S z$, or $z = x/H_S$. For the transfer function H_S of the combination of valve and linear motor we can write: $H_S = K_S/(s + b_1 s^2 + b_2 s^3 + \ldots)$. Here K_S is the time-independent part of the function (referred to as the static gain), b_1 and b_2 are constants characterizing the dynamics of the system. The

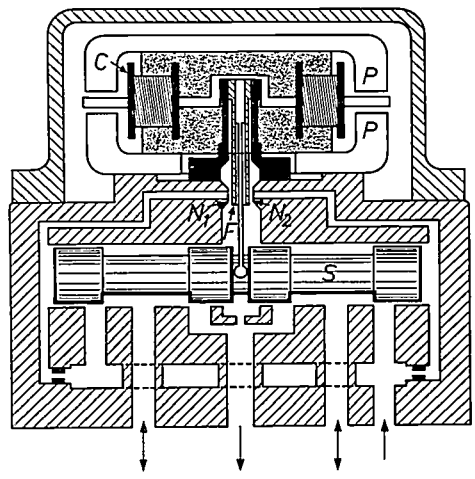


Fig. 12. Cross-section of the 'MOOG' servo valve. *P* pole pieces of magnetic circuit. The armature between the pole pieces is tilted by a control current through the coils *C*. The armature moves a vane *F*, which controls the outflow resistances of the two nozzles N_1 and N_2. Both nozzles are connected via fixed flow resistances with the oil supply, so that a movement of the vane causes a pressure difference across the spool *S*, which thus moves accordingly. The spool regulates the oil feed to the two chambers of the linear motor. A more linear relation between electric control current and oil outflow is obtained by including mechanical feedback between the spool and the vane *F*.

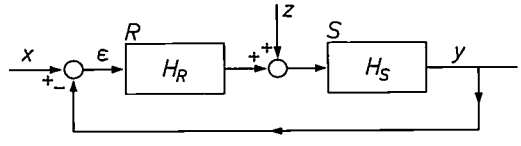


Fig. 13. Calculation of the forward-drive signal.

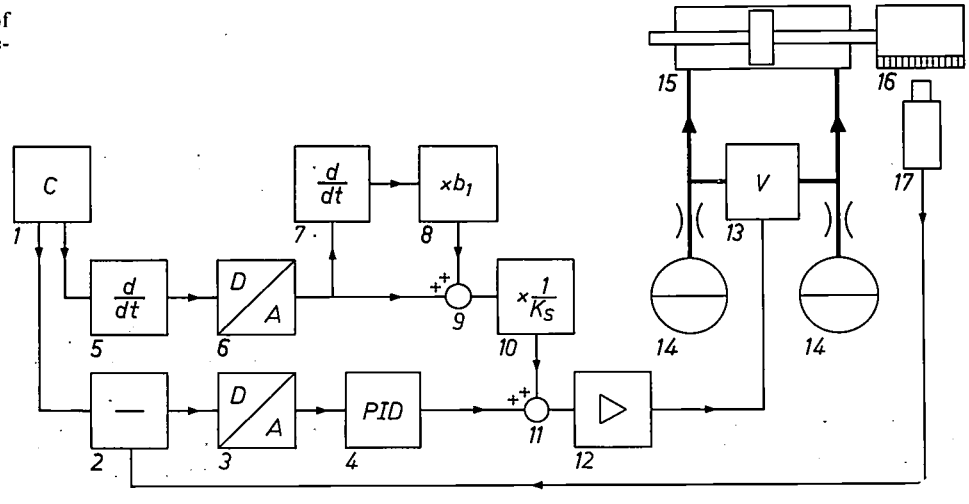


Fig. 14. Complete diagram of the control loop for the displacement of the slide.
1 Control computer
2 Digital subtraction circuit
3 Digital-to-analog converter
4 PID controller
5 Digital differentiator
6 Digital-to-analog converter
7 Analog differentiator
8 Multiplier (factor b_1)
9 Summation point
10 Multiplier (factor $1/K_S$)
11 Summation point
12 Amplifier
13 Servo valve
14 Dampers
15 Linear motor
16 Slide
17 Digital measuring instruments

symbol s is the differential operator d/dt; the equation is therefore in fact a differential equation describing the dynamics of the system. Inserting the relation just found for z gives:

$$z = (1/K_S)(s + b_1 s^2 + b_2 s^3 + \ldots)x = \\ = (1/K_S)(dx/dt + b_1 d^2x/dt^2 + b_2 d^3x/dt^3 + \ldots).$$

Since we only have signals available for speed and acceleration we have to limit this expression to $z = (1/K_S)(v + b_1 a)$. This gives the operation to be performed on the signals for v and a to obtain the desired feed-forward signal.

The complete diagram of the control loop is given in *fig. 14*. Only the position of the carriage is measured and fed back. Unfortunately in our case no sufficiently accurate transducers are available for the speed and acceleration, since they are linear and not rotational quantities; otherwise a further improvement in the control loop could have been obtained by feeding back these measured values.

Damping in the hydraulic motors is minimal, as is usually the case in mechanical systems. To improve the stability of the control loop we therefore had to introduce extra damping. This took the form of two small reservoirs, each connected by a laminar restrictor to one of the two motor compartments [7]. (These damping reservoirs are each equivalent to RC smoothing filters.) The stability of the control loop is most critical when the piston is in the middle of the cylinder. If the measures taken to meet this situation are adequate, the loop will be sufficiently stable for all other positions of the piston.

Just why this middle position of the motor piston is the most critical situation can be understood as follows. The carriage connected to the piston may be regarded as a solid mass supported on two springs formed by the oil columns on both sides of the piston. The stiffness of these springs is inversely proportional to the lengths of the oil columns. The natural frequency of the spring-mass system, which is mainly determined by the stiffer spring, is now at its lowest value when both springs are equal. At these low frequencies the mechanical damping is also very low and consequently the system is then in its most critical situation.

Root locus of the control loop

The stability of the control loop can be derived from the behaviour of the poles of the transfer function $H(s)$. The poles of H are the roots of the characteristic equation of the reduced differential equation for the control loop [8]. This reduced differential equation describes the behaviour of the control loop when there is no input signal and the loop is thus left to itself. The poles or roots are plotted in the complex plane as curves with the static gain K as parameter. For the loop to be stable, all poles must have a negative real part, since a pole λ gives rise to a term $e^{\lambda t}$ in the solution of the reduced differential equation for the control loop, and this term will only represent a damped signal if it has a negative real part.

[7] J. J. 't Mannetje, Stabilizing networks for hydraulic motors, Control Engng. 21, No. 6, 55-58, June 1974.

[8] A more detailed discussion is given in J.-C. Gille, M. J. Pélegrin and P. Decaulne, Feedback control systems, McGraw-Hill, New York 1959, chapter 14.

Fig. 15 gives the root locus of the slide-control loop for the most critical situation, i.e. when the motor piston is in the middle position. The K values at which the curves cut the real axis are indicated at the points of intersection. To maintain stability, these values must under no circumstances be exceeded. In practice operation is normally confined to the area bounded by two lines at 45° to the negative real axis. Since the damping is exactly critical on these lines, the damping in the bounded area is always to a greater or lesser extent over-critical.

A final specification is that the loop gain must be identical in the control loops for both slides, to ensure that oblique lines are in fact drawn with the required slope. The performance of the control loops is clearly demonstrated by the perfection with which the test plate in *fig. 16* has been drawn.

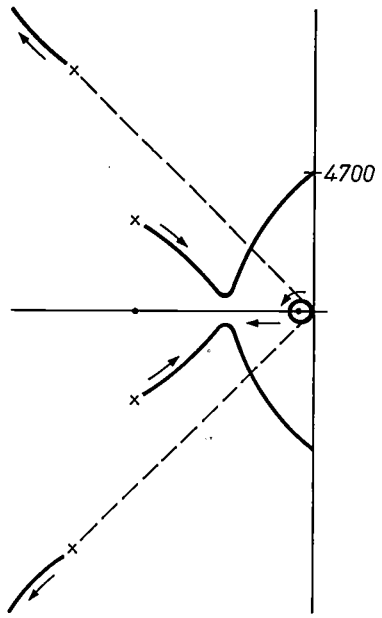


Fig. 15. Root locus of the control loop shown in fig. 14.

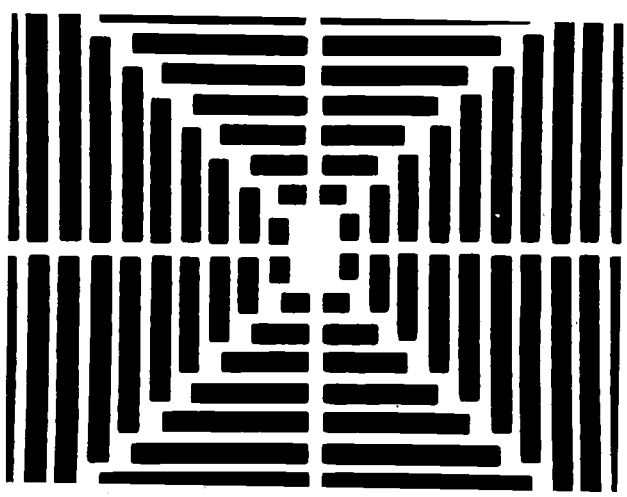


Fig. 16. Test pattern drawn with the Opthycograph. The smallest blocks in the centre measure 10 by 15 μm. Because of the very regular pattern chosen, slight deviations from this regularity would be very noticeable.

The computer control

Regarded as a numerically controlled machine, the Opthycograph is a five-axis contour-controlled machine. The five axes refer to the fact that five quantities can be varied independently of each other. These quantities are: the x and the y coordinate, and the length, width and angular position of the slit. Contour control implies that it is not only the beginning and end points of a line that are well defined but also a number of points in between (*fig. 17*). This number is chosen such that the deviation between the path which the light spot is desired to follow and the path that it actually follows is always less than $\frac{1}{4}$ μm. Since the drive always has some inertia, the actual path will deviate even less from the desired one.

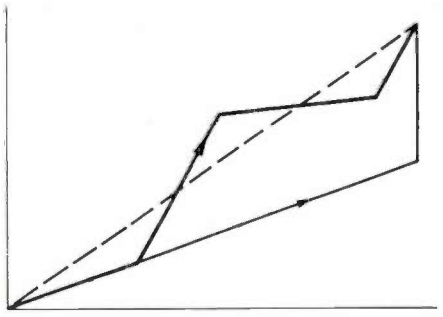


Fig. 17. Point-to-point control and contour control. The thinly drawn line gives the track followed in point-to-point control, the thick line indicates the track followed in contour control. The dashed line gives the ideal track between the start and end points.

In designing the Opthycograph we chose the Philips P9201 computer as the principal unit of the control system. This enabled us to program the great majority of control details in the software and to restrict the actual control hardware to a minimum. As a result the control is flexible, changes in the control data requiring only a modification in the software. Since the hardware does not have to be modified, the machines can be used for other work with the still unmodified program while the changes are being prepared.

Input of the control commands is effected on punched tape or magnetic tape. The commands can be in a readable code, in which each movement of the machine is indicated by a group of characters, called a block, terminated by the 'end-of-block character' (EOB). A command can thus read:

X 1000 Y — 120 M3 T120 (EOB),

which means: move the x-slide 1000 μm in the positive direction and simultaneously the y-slide 120 μm in the negative direction, with light on (M3) and a line width of 120 μm (T120).

An advantage of this readable code, which has been standardized for numerically controlled machines, is that a workpiece can be set out or corrected by hand on a tape punch. Usually, however, as was mentioned at the beginning, workpiece tapes are nowadays directly delivered by a large computer used in arranging the elements of an IC. For general geometries the program for this computer can be written in APT [9], a programming language for the numerical control of machine tools; for integrated circuits a special programming system has been designed. The total commands for an IC with 80 elements (transistors, capacitors and resistors) now include some 2×10^3 characters (60 metres of punched tape), but there are even larger series of commands, e.g. 10^4 characters for one mask.

As mentioned earlier, a straight line can be indicated in one command. An arc can also be given as one command, though it is approximated to within the specified tolerances by straight line segments. For other curved lines this approximation by straight line segments must be incorporated in the control program. In practice this is done by means of the APT language.

The only specified parts of the tracks to be drawn are the centre line and the start and end points, with the required width. This means that there will be gaps at places where two of these tracks meet at an angle, and therefore instructions to fill these up are included in the computer software (*fig. 18*). Bends in a track can be automatically rounded, making the track parabolic on both sides of the point of intersection. This has a dual advantage: the necessary calculations, as we shall see, can be performed very simply and therefore quickly by the computer, and the movement along a parabolic path is very smooth because the accelerations are constant.

Internal communication

The speed and organization of the internal communication between the main parts of the Opthycograph — control desk, control computer and light plotter — is determined by the sampling frequency, i.e. the frequency at which the computer accepts and delivers control data. The sampling frequency, which should be about ten times the bandwidth of a control loop for

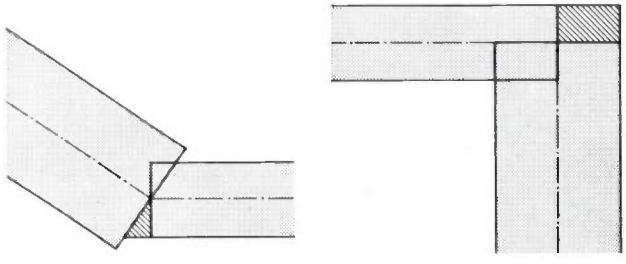


Fig. 18. Filling gaps. When lines of the desired width are drawn between start and end points, only the grey areas are produced. Separate instructions in the computer program ensure that the shaded areas are filled up.

satisfactory operation, was taken as 1500 Hz. For communication between the computer and the other parts of the machine there are four parallel channels of 16 bits each. One channel is used for the displacements of the x- and y-slides, a second channel for controlling the slit, and the other two channels are for communication with the control desk. The capacity of these communication channels is sufficient for the operating speed of the Opthycograph. Numerical data will be found in *Table III*.

To ensure good communication between the computer and the other parts of the machine particular attention has to be paid to certain aspects of the electronics. Since the Opthycograph is set up in a dust-free room, the distance between the machine and the computer is greater than is usual between a computer and its peripherals. The cable impedance and the level of the transmitted signals are given values compatible with this situation to minimize the effects of interference, cross-talk between cable wires and severe signal distortion. The very low-level output signals of the computer are first amplified in power amplifiers to make them suitable for operating the various control and drive units of the machine. It was also necessary to build a control circuit for the stepping motor that rotates the Péchan prism to alter the angular position of the slit. Since virtually no overshoot can be tolerated here, an 'optimal switching' system was chosen for this purpose. In this kind of control each accelerating pulse is always followed by a retarding pulse of equal magnitude, giving the fastest possible drive for the stepping motor with hardly any overshoot [10].

Software

The permanent software in the computer memory provides for the processing of the input data and delivers it in a form suitable for controlling the Opthycograph. The data is at the same time checked for errors and incompatibility, and the correction, fed in on separate tapes, are taken into account. The required path velocities and accelerations are calculated, both for controlling the slit length — exposure — and for controlling the carriage movements. Finally, the angles for determining the angular orientation of the slit are calculated and the data is determined for rounding off corners and filling up gaps where two lines meet.

All these operations of the computer are interrupted 1500 times per second, i.e. every 700 microseconds, in favour of the sampling program responsible for direct communication with the Opthycograph proper. This sampling program usually takes up about 350 μs, half of the available 700 μs, but the extra time has to be provided since the sampling program can take 660 μs sometimes.

Table III. Independently controllable quantities, their maximum and minimum values and the speed at which they can be changed. Lengths are measured using the smallest step of 0.5 μm as a unit; the unit of time used is the sampling time of 700 μs.

Quantity	Min. - max. value	Fastest possible change
Speed in x- and y-directions	0-15 steps/sampling	1 step/sampling per two samplings
Slit length		± 1 step per sampling
Slit width	2-1900 μm in 200 steps	one step/ 8 samplings
Slit rotation		± 1°/sampling

A computer operates sequentially, that is to say the instructions are processed one after the other, so that the computer only handles one program at a time. Apart from the sampling program, which interrupts the drawing operations every 700 μs, all other programs are started in their turn by a monitor program. In this way each part of a program gets a chance to perform its function once in every cycle. If a function cannot be continued because the available time is used up, or perhaps because there is not enough room left in the buffer store to which the results are temporarily transferred, the program is interrupted and resumed during the next cycle.

Calculating methods

In conclusion we shall discuss some calculating methods used in the programs mentioned above. The methods are designed to permit the calculations to be carried out quickly with the small computer. A fairly generous period of time is available for the preparatory operations that are required for translating the control commands into a form that can be used directly for controlling the Opthycograph. After interruption by the sampling program the operations can be continued. Nevertheless it is always necessary to find the simplest possible algorithms, and procedures such as series expansions are too time-consuming. The requirements for the sampling program are more difficult since there is never more than 700 μs per calculation available. This program is therefore only used for those calculations that could not be carried out earlier owing to the lack of certain data. The time limit even rules out simple multiplications of two words of 16 bits, since the P9201 is not equipped with a 'hardware multiplier'.

[9] J. Vlietstra, Philips tech. Rev. 28, 329, 1967.

[10] 'Optimal switching' is also dealt with in J.-C. Gille, M. J. Pélegrin and P. Decaulne, Feedback control systems, McGraw-Hill, New York 1959, p. 443, and in J. E. Gibson, Nonlinear automatic control, McGraw-Hill, New York 1963, chapter 10. The principle is described, without mentioning the name, by P. J. M. Janssen and P. E. Day, Philips tech. Rev. 33, 190, 1973 (No. 7), and by J. Crucq, Philips tech. Rev. 34, 106, 1974 (No. 4).

The methods of calculation must therefore be extremely efficient in use of computer time and storage capacity. We have achieved this efficiency by confining the arithmetic operations primarily to adding and subtracting, and keeping operations such as multiplication to a minimum. If a multiplication is unavoidable, we take a positive or negative power of two as the multiplier. This reduces the multiplication to shifting numbers in a binary register, which is just as simple and fast an operation as adding.

Dividing a straight line into segments

A line is divided into segments to enable further calculations to be performed within one word length (16 bits). Only the lengths of the x- and y-components of the line are known, the length of the line itself is not. We look for the larger of the two components and divide it by 2^n, giving n a magnitude such that the quotient lies between 0.5 and 1 mm; the length of the line segment is then between 0.5 and $\sqrt{2}$ mm. The other component is now divided by 2^n as well. The remainders of these divisions are then stored and added again in the processing of a line segment.

Parabolic rounding-off

The start and end points of the rounding-off are taken at a distance of 100 μm from the point of intersection of two straight paths (*fig. 19*). The top of the rounding-off curve is chosen halfway between this intersection and the line connecting the start and end points. Taking the line segments as vectors, as indicated in the figure, we can then express the chosen position of the top by $\boldsymbol{A}_4 = \frac{1}{4}(\boldsymbol{A}_2 - \boldsymbol{A}_1)$. We now divide the curve between the start and the vertex into N parts and fix the points of division by the relation

$$\boldsymbol{P}_n = (n/N)\,[\boldsymbol{A}_1 + (n/N)\boldsymbol{A}_4].$$

These points then lie exactly on a parabola, as can be seen if we resolve the vectors $\boldsymbol{P}_n$ into components perpendicular to and parallel with $\boldsymbol{A}_4$. The expressions then found correspond to the parametric form of the equation for the parabolic projectile path.

In drawing the rounding-off curves we approximate to this parabola by connecting up the dividing points with straight line segments. The segment between $\boldsymbol{P}_{n-1}$ and $\boldsymbol{P}_n$ (the first difference) is given by

$$\frac{1}{N}\left(\boldsymbol{A}_1 + \frac{2n-1}{N}\,\boldsymbol{A}_4\right),$$

while the second difference, $\Delta^2\boldsymbol{P}_n = \Delta\boldsymbol{P}_n - \Delta\boldsymbol{P}_{n-1}$, is given by $\Delta^2\boldsymbol{P}_n = 2\,\boldsymbol{A}_4/N^2$. This second difference is thus constant and can be designated as $\Delta^2\boldsymbol{P}$. We can now express the length of the successive line segments as $\Delta\boldsymbol{P}_n = \Delta\boldsymbol{P}_{n-1} + \Delta^2\boldsymbol{P}$.

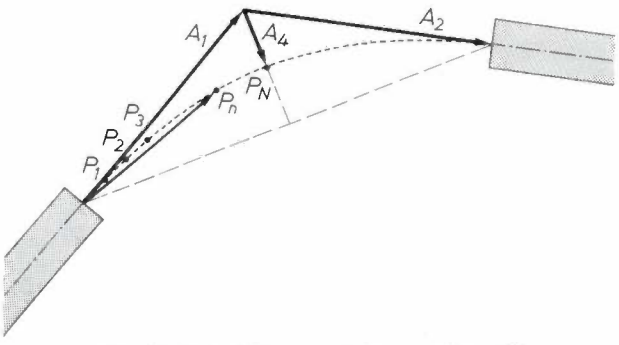


Fig. 19. Calculating parabolic rounding off.

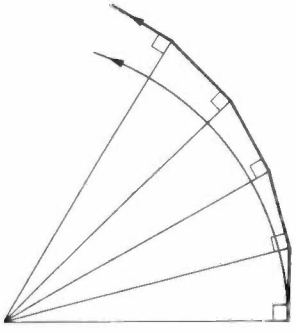


Fig. 20. Circular interpolation. The circle is approximated by a polygon. In the approximations used, each line segment is perpendicular to the radius of the circle at the beginning of this segment, so that the polygon approximates to a spiral and not a circle. After a number of steps a correction therefore has to be applied.

The dividing points on the right-hand half of the parabola are found by starting from

$$\boldsymbol{P}_n = (n/N)\,[\boldsymbol{A}_2 - (n/N)\boldsymbol{A}_4].$$

Expressions of the same form are found for the x- and y-components of the line segments. If we choose a power of two for N, as mentioned earlier, e.g. 8, the calculations are easily carried out.

Calculation of angles and path lengths

To calculate the angle at which a track runs, which is necessary, for example, for controlling the slit orientation, we start from the ratio between the smallest increment a and the largest increment b. A table with a/b as input and 64 values for this quotient makes it possible to determine angles accurately to within 25′ or even better. This is good enough for determining the orientation of the slit. The path length is calculated from $\sqrt{a^2 + b^2} = b\sqrt{a^2/b^2 + 1}$; the value of $\sqrt{a^2/b^2 + 1}$ is included in the table.

For circular interpolation it is necessary to be able to calculate angles to an accuracy of 2′. This calculation is carried out by means of interpolation between two successive values from the table.

Linear interpolation

In the preparatory processing of the control commands the data for the displacement of a slide is calculated as if the tracing speed at the beginning of the

operation were one unit per sampling. The sampling program now calculates the displacement actually necessary from the speed communicated to the slide control. The unavoidable multiplication is very simple in this case. The two factors have a small value since, as just described, the path has previously been divided into segments with a length of no more than $\sqrt{2}$ mm.

Circular interpolation

To draw circles or connect two points by a circular arc, we use a circular interpolation. In this procedure a circle is approximated by a polygon with a very large number of sides. The coordinates of each successive vertex are calculated from those of the previous point, using recurrent equations given in parameter form. The parameter is the angle $\Delta\phi$ between two successive points. Since this angle is very small, we approximate to the sine and cosine functions by the first term of the power series for these functions. This approximation means that each line segment is perpendicular to the circle radius at the beginning of the segment (*fig. 20*). Consequently the path drawn is not a circle but a spiral, and a correction therefore has to be made whenever the error becomes too great.

The instant at which the correction should be applied can be calculated as follows. After n steps the radius R_0 with which we started has increased to R_n. When the same distance A is travelled in each sampling, we can write $R_n^2 = R_0^2 + nA^2$. We now wish to correct the radius when the deviation amounts to a quarter of a unit. At that moment we therefore have

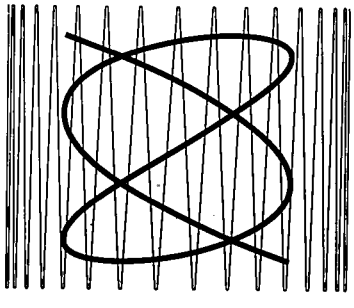
$$R_n^2 = (R_0 + \tfrac{1}{4})^2 = R_0^2 + \tfrac{1}{2} R_0 + \tfrac{1}{16}.$$

We shall not commit a serious error if, instead of applying the correction exactly at this moment, we do so when the second terms in the two expressions for R_n^2 are identical, that is when $\frac{1}{2} R_0 = nA^2$.

If the path travelled per sampling is not constant, we must then substitute ΣA_j^2 for nA^2. This sum can be calculated from a table in which the value of A_j^2 can be found for every A_j. This has again been done to avoid the time-consuming process of calculating during the execution of a command.

To end this article we should like to mention the valuable contributions made to the completion of the Opthycograph by A. C. J. M. van Asten, V. A. van de Hulst, A. C. Jacobs, A. M. T. H. Jentjens, G. C. van de Looy, W. van der Meulen and J. Rienks.

Summary. With the increasing complexity of integrated circuits (ICs) it is becoming an impractical proposition to make photomasks for these ICs by the usual cut-and-strip technique. The precision instrument described here provides a way out of this difficulty. The machine, called the Opthycograph, draws with a moving light beam on a photographic plate. The optical system and the photographic plate move on hydraulically driven slides. The slides are mounted on hydrostatic bearings. The whole system is controlled by a small computer. The repetition accuracy of the drawing is $\pm$ 0.5 μm; the largest drawing field is 200×200 mm. The article gives a general description of the machine and a more detailed description of the drive-control and computer-control systems. Finally, a number of calculating methods are discussed, showing how the computer input containing the specification of a mask geometry is converted into the control commands for the drawing machine.



Philips tech. Rev. 34, 270-273, 1974, No. 10

Investigation of microchannel plates by scanning electron microscopy

Microchannel plates, which consist of a compact bundle or stack of glass electron-multiplier channels, are being increasingly used in image intensifiers, radiation-detector arrays, etc. [1]. The inside diameter of the channels varies at the present time from 100 μm down to 12.5 μm, so that a very powerful microscope is required for examining them optically. A microscope eminently suitable for this purpose is the scanning electron microscope [2]. This not only gives the required magnification (100 000 times) and a high resolution (15 nm) but also enables a study of the secondary emission from the inside wall of the glass channels to be made. This secondary emission is responsible for the electron-multiplication process through the channels. We have attempted in our investigation to make the best use of the facilities offered by the scanning electron microscope for this purpose.

The microchannel plates examined were discs with a thickness of 0.66 mm and a diameter of 22 mm. The channel diameter was 16 μm and the channel pitch 20 μm. The electron gain of the channels was 10^4 at a voltage of 1 kV.

For examination the plates are mounted in a holder specially designed to permit electrical contact with each of the electrodes vacuum-deposited on the flat faces of the channel plate. The holder does not present an obstacle to the escaping electrons on their way to the electron detector.

The microscope scanning beam, which has a focal-point diameter of the order of 0.01 μm, enters the channels as it scans the plate and releases secondary electrons from the channel walls. Which part of the wall is 'seen' by the beam depends on the orientation of the plate in relation to the beam axis.

A variety of information can be obtained by varying the choice of the potential difference between the electrodes. If both electrodes are earthed, the secondary electrons generated in the channels play no part in the picture obtained. If the input face scanned by the electron beam is earthed and the output face negatively biased (e.g. to —350 V), the channel plate does not then operate in the normal mode but in the reverse mode: the secondary electrons released by the beam and subsequently multiplied emerge from the input face as illustrated in *fig. 1a*. For operation in the normal forward mode the output face is earthed and the input face is biased to the normal operating voltage of about —800 V (fig. 1*b*).

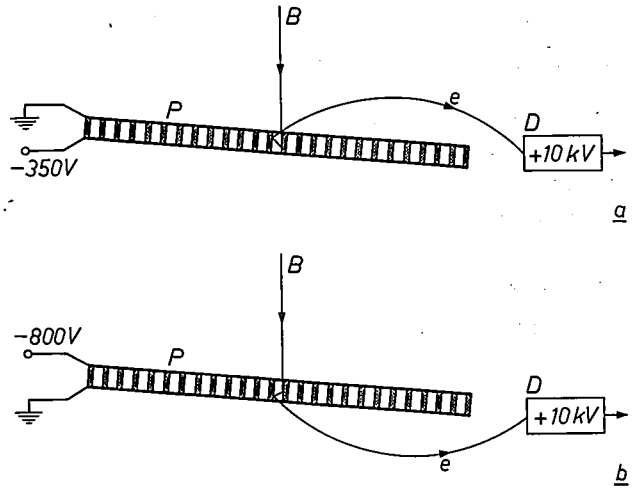


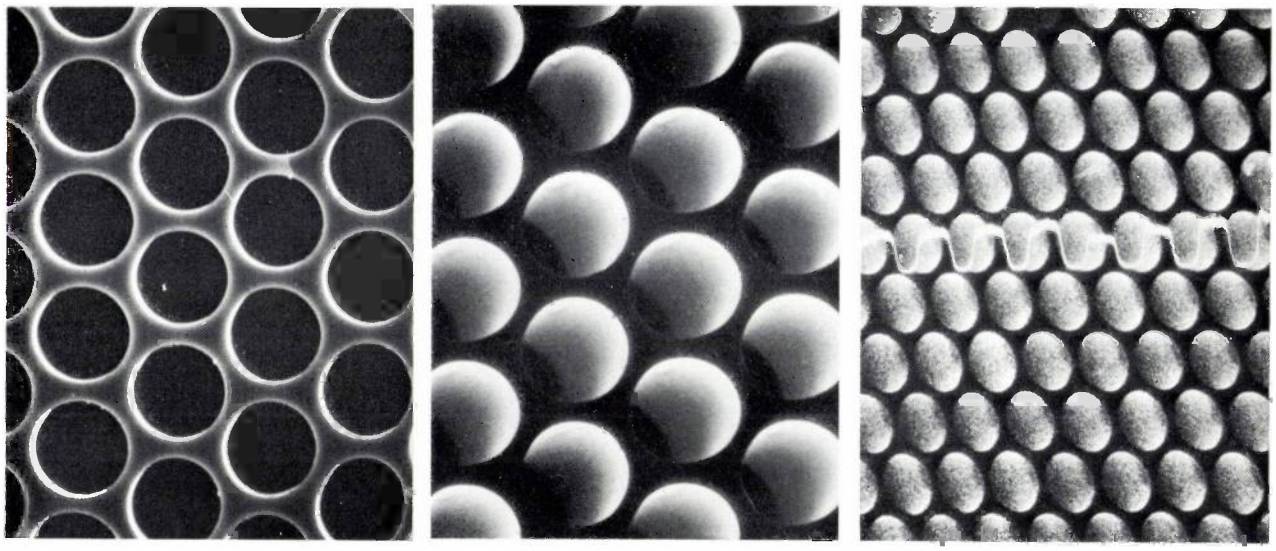
Fig. 1. Schematic arrangement of two methods of examining a microchannel plate by scanning electron microscopy. *P* microchannel plate. *B* primary electron beam. *D* secondary-electron detector. *a*) The electrode on the output face is at a negative potential with respect to the input face. Secondary electrons released in the channels emerge from the input face. *b*) The electrode on the input face is at a negative potential with respect to the output face. The secondary electrons now emerge from the channels, as in normal operation, at the output face.

Fig. 2 shows scanning-electron photomicrographs illustrating each of these three modes. Fig. 2*a* relates to the case where both electrodes of the channel plate were earthed. The electron micrograph shows the perfect circular cross-section of the channels and their regular geometrical structure. Fig. 2*b* shows the same channel plate, but with the potential of the output-face electrode at —300 V (reverse mode). The gradation of the image of a channel depends on the electron gain. By varying the focusing it is possible to obtain sharp images of zones of a channel wall, so that features such as surface irregularities can be observed.

When the electrode on the input face is negatively biased, as it is in the normal mode of the channel plate, the electrons released from the channel wall by the scanning beam are multiplied in the forward direction and collected behind the output face by the detector.

[1] G. Eschard and B. W. Manley, Acta Electronica **14**, 19, 1971.
D. Washington, V. Duchenois, R. Polaert and R. M. Beasley, Acta Electronica **14**, 201, 1971.
J. Graf and R. Polaert, Acta Electronica **16**, 11, 1973.

[2] O. C. Wells, Rev. sci. Instr. **40**, 1246, 1969.
G. Eschard and R. Polaert, The production of electron-multiplier channel plates, Philips tech. Rev. **30**, 252-255, 1969.
A. M. Tyutikov, V. K. Kozyrev, V. P. Puzanova and Yu. V. Chentsov, Optical Technology **39**, no. 1, 1972.



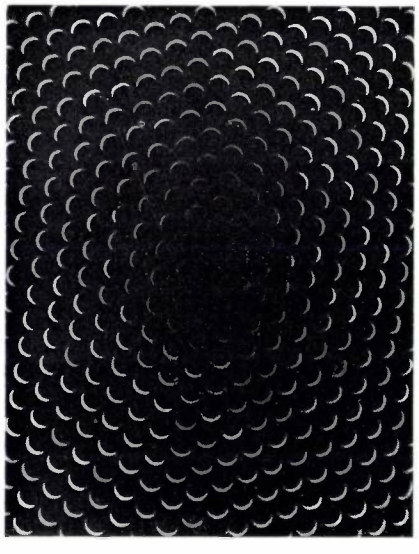
a *b* *c*

Fig. 2. Electron photomicrographs of a microchannel plate, taken with a scanning electron microscope. *a)* Both electrodes earthed. *b)* Input face earthed, output face at −300 V. *c)* Input face at −800 V, output face earthed. The photograph also shows the video signal of a line scan.

The image then displayed on the monitor is a measure of the electron gain in each channel, serving as a useful basis for assessing the homogeneity of the gain obtained from the channel plate (fig. 2*c*). When a line is scanned it is possible to display the video signal separately; the amplitude of this signal gives a quantitative comparison of the electron gain of all the channels on this line.

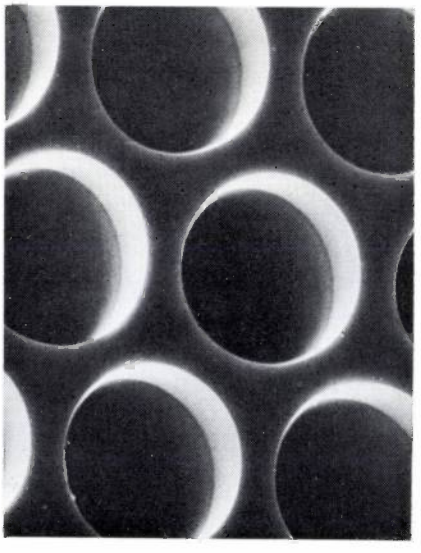
By rotating the plate the channels can be oriented in such a way that the axes of the channels are parallel to the axis of the scanning beam. The beam does not then collide with the wall and thus passes through the channel without generating any secondary electrons; this situation corresponds to the central dark zone in *fig. 3*. This central zone is surrounded by a region where the beam enters the channels suffiently obliquely to strike the wall, giving a concentric pattern of 'crescent moons'; the regularity of these crescents is a good test of whether the channels are parallel. Starting from the centre point, where the beam and channels are exactly parallel, it is easy to calculate the angle between two axes elsewhere on the plate. This permits a measurement of the penetration depth of the metallized electrodes inside the channel (*fig. 4*).

If the electron gain is high (e.g. 10^4) the scanning beam should be kept to a low intensity in order to avoid saturation of the secondary electron current in the channels. In that case (see fig. 2*c*) the picture has a speckled appearance: an individual white spot is the amplified image of a single electron event at the entrance of the channel.



←

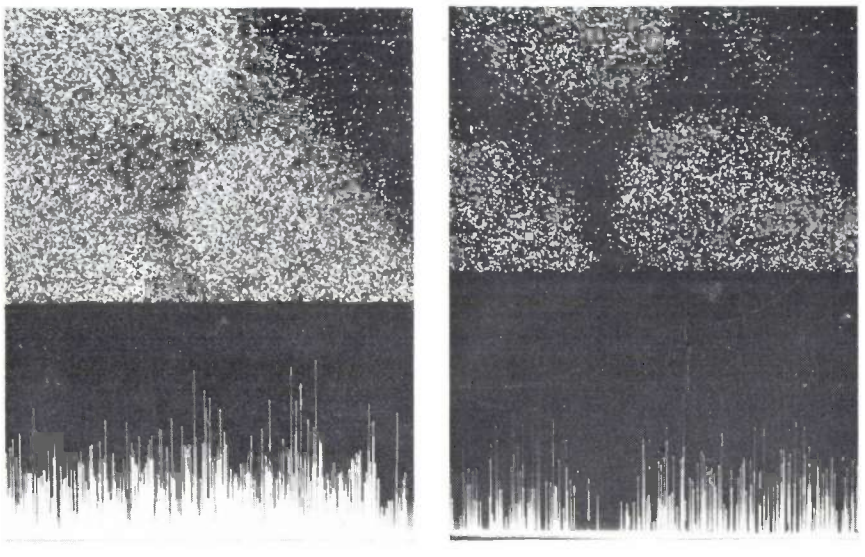
Fig. 3. Microchannel plate with the central part of the photograph perpendicular to the scanning beam. The electron beam passes through the channels here without striking the wall.



→

Fig. 4. Illustration of the penetration of the vacuum-deposited electrode material in the channels, where use is made of the difference in secondary emission from the electrode material and from the glass channel wall. The contrast also depends on the beam intensity. The depth of penetration can be determined from the orientation of the plate with respect to the primary beam (see fig. 3).

The image formation is due not only to the primary electrons directly entering a channel but also, to some extent, to backscattered and secondary electrons from the input face, that is to say from the cross-sections of the intermediate walls, which are redistributed to other channels. What effect do these redistributed electrons have on the image quality of a channel plate examined with a scanning electron microscope? This is a particularly important point in those image tubes in which the field in front of the channel plate is such that redistributed electrons are returned to it, e.g. proximity-focused tubes. A configuration of this type can be simulated in the microscope by means of a fine mesh (100 microns) placed in close proximity (300 microns) to the input face, the mesh being given a negative potential (—90 V) with respect to the input face. Most of the redistributed electrons are then re-injected into one of the neighbouring channels where they are multiplied.



a *b*

Fig. 5. Electron photomicrographs illustrating the effect of the redistributed electrons from the input face. The energy of the primary electrons is 2.1 keV. *a)* The fine mesh in front of the plate has a potential of —90 V with respect to the input face. *b)* Potential of the mesh +90 V with respect to the input face. The video signal for a line scan can be seen below the photomicrographs. This signal gives a quantitative picture of the electron gain, which depends on the site where the primary beam strikes the channel plate.

We determine the contribution of the redistributed electrons from the input face by counting the number of electron pulses collected during a particular time from a large number of sites of a given area. At a primary electron energy of 2.1 keV the contribution was found to be no less than 17% when the mesh was at a potential of —90 V, but less than 1% when the potential was reversed. This means that the total *effective* open area of the channel plate in this experiment is 63% + 17% = 80% (the 63% corresponds to the geometrical open area of the channels). *Fig. 5* shows two cases with very large magnification. All the white spots in the rather darker zones in fig. 5*a* were due to electrons coming from the input face. From the spectrum of the detector signal during line scanning (lower part of fig. 5*a*) it is hardly possible to determine where the scanning beam passes a channel wall. Fig. 5*b* shows the situation in which the electrons redistributed from the input face are nearly all collected by the positive mesh and make no further contribution to the image. The walls between the channels can now be distinguished much more clearly as dark zones.

The range of the redistributed electrons with a negatively biased mesh can be determined by screening off part of the channel plate at the output face. The screened-off channels are then non-active. If electron impacts are still displayed when the corresponding part of the input face is scanned, the display can only originate from electrons that have been released from the wall cross-section and have then been multiplied in unscreened neighbouring channels. This is illustrated in *fig. 6* with three different magnifications. These experiments show that the mean range of these redistributed electrons is about 150 μm. The conditions used in examples illustrated differ from those which occur in a real image tube such as a proximity-focused tube. Here the primary energy of the beam is 0.3 keV and the field in front of the channel plate 1 kV/mm;

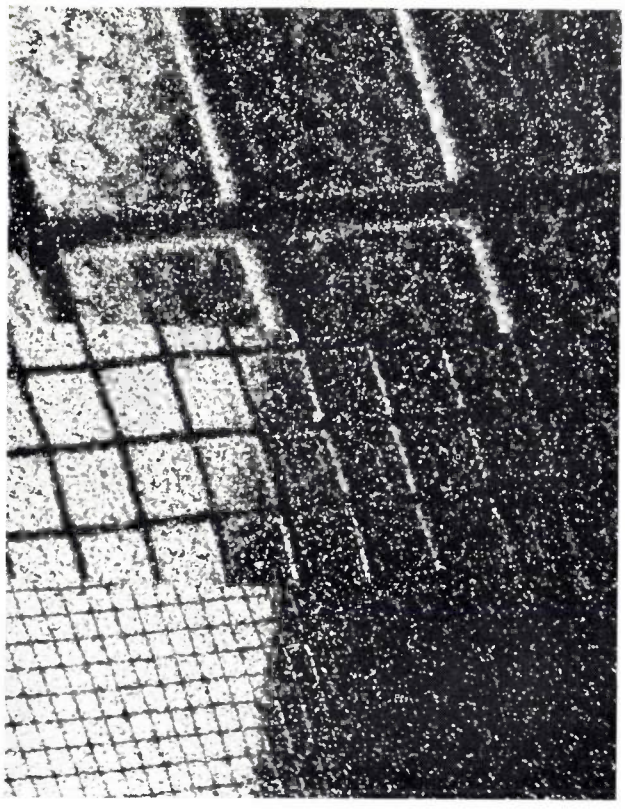


Fig. 6. Scanning electron photomicrograph, with three different magnifications, from which the mean range of the redistributed electrons can be determined. The lattice width of the mesh shown here is 100 microns. The mesh is at a potential of −90 V. The dark zone on the right is a screened part of the channel plate. The electrons detected during the scanning of this part of the plate were released from the input face, then backscattered and amplified in unscreened channels.

it is these factors which will determine the range of the redistributed electrons. The studies will be continued simulating more closely the conditions found in a real tube.

The investigation has shown that the scanning electron microscope is an eminently useful means of analysing such diverse aspects of the quality of microchannel plates as electron-gain homogeneity, local saturation effects, the state of the channel walls, quantum efficiency and the effect of the secondary electrons on the contrast.

R. Polaert
J. Rodière

R. Polaert, ingénieur H. E. I., Dr. Ing., and J. Rodière are with Laboratoires d'Electronique et de Physique Appliquée (LEP), Limeil-Brévannes (Val-de-Marne), France.

Philips tech. Rev. 34, 274-287, 1974, No. 10

The Phototitus optical converter

F. Dumont, J. P. Hazan and D. Rossier

In the last ten years there has been increasing interest in the optical processing of images and other data. One result of this has been an increased demand for devices that can fulfil the function of real-time re-usable film. In recent years several such 'image converters' have been developed; 'Phototitus' is one of them. Phototitus will store images, add or subtract them, and process them in other ways. An important feature that distinguishes Phototitus from most other image converters, and could well be of interest in the field of coherent-optical processing, is that a stored image can be read out with coherent light without destroying the coherence of the light. In Phototitus the read-out light is very efficiently used and only moderate voltages are required for operation. The attainable contrast, image resolution and write-in sensitivity compare well with those of other devices.

If a thousand pencils are scattered across a table, and one is then removed, an observer who sees the scene 'before' and 'afterwards' would find it hard to tell the difference. Even if pictures of the two scenes are presented to him for comparison, it might take him quite a while to find out that they differed by one pencil. But if a 'negative' of the second picture is superimposed on the positive of the first, the difference shows up immediately.

This 'picture subtraction' is an example of 'image processing'. An essential part of the processing is the storage of an image until it has been dealt with. Images could be processed by using photographic film. However, while film is ideal for storage, such processing is tedious, complicated and often unsatisfactory.

Phototitus [1] is a device that can subtract images — or perform other operations on them — instantaneously. Its multilayered structure is shown schematically in *fig. 1*. It consists of a single crystal C of deuterated potassium dihydrogen phosphate (DKDP), a dielectric mirror M, a layer L of amorphous selenium and two conducting transparent electrodes A_1 and A_2. The selenium is photoconductive. If an optical image is projected on it, electron-hole pairs are created near A_1. If A_1 is made positive with respect to A_2, some of the holes drift towards the interface with the dielectric mirror M, where they are stored, presumably in surface traps. At any point of the interface, the accumulated charge is proportional to the intensity of the incident light. The optical image is thus converted into a latent charge image, as in xerography. When a new image is projected, the corresponding latent image adds to the first. If an image is projected while A_1 is *negative* with respect to A_2, electrons instead of holes drift towards the interface, and the corresponding charge pattern subtracts from the latent image already there.

The DKDP plate allows the latent image to be read out. This operation depends on the Pockels effect of DKDP: linearly polarized light passing through the crystal emerges elliptically polarized if there is an

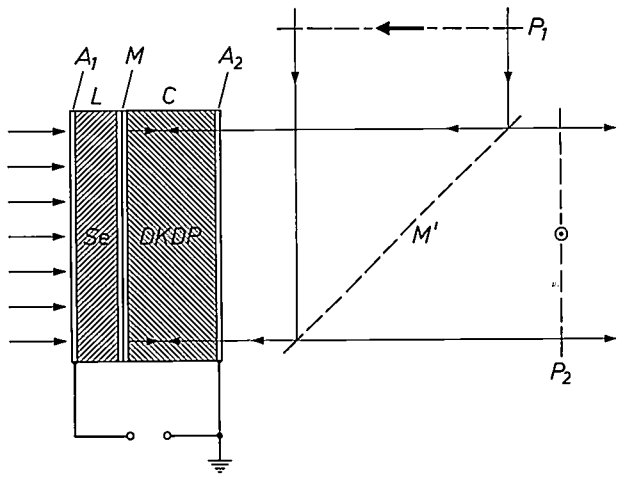


Fig. 1. Principles of Phototitus. L photoconducting layer of amorphous selenium. C single crystal of DKDP. M dielectric mirror. A_1, A_2 conducting transparent electrodes. P_1, P_2 crossed polarizers. M' semitransparent mirror. An optical image projected on L from the left produces a latent charge image, which is stored at the interface with M. The latent image is positive, consisting of holes if A_1 was positive during the projection, and negative, consisting of electrons if A_1 was negative. The latent image polarizes the DKDP. Owing to the Pockels effect in the DKDP the beam of polarized light is modulated in accordance with the latent image after it has passed twice through the DKDP and had also passed P_2.

Dr Ing. F. Dumont, Dr J. P. Hazan and Dr D. Rossier are with Laboratoires d'Electronique et de Physique Appliquée (LEP), Limeil-Brévannes (Val-de-Marne), France.

electric field across it. The larger the field, the larger the deviation from linear polarization, and therefore the larger the amount of light transmitted by a crossed analyser. Thus the light beam in fig. 1, after having passed the analyser P_2, is modulated by the latent image: bright patterns correspond to high-field areas in the DKDP plate, i.e. areas of high charge density in the latent image.

Phototitus is derived from the TITUS tube [2]. In this tube, the main element is also a plate of DKDP material, but here the image is written into the device electronically, by means of a swept electron beam and a video signal across the DKDP plate. In both cases the DKDP plate is cooled to just above the Curie temperature (about —50 °C). In this way a large Pockels effect is obtained: the effect is proportional to the dielectric constant ε_c along the optical axis, which increases strongly near the Curie temperature.

Phototitus belongs to a large family of devices developed recently for image storage and processing, all based on the idea of combining a photoconductor with an electro-optic material. In a later section we shall shortly compare Phototitus with other devices of this family. Here we want to stress one feature that it shares with only a few of these: it allows a stored image to be read out with a laser beam, without destroying the coherence of the laser light. It can thus convert print on paper into images that can be analysed by coherent light. This may open the way for *real-time* coherent optical processing — such as the recognition of patterns or characters using holographic filters — of images printed on paper. This feature is present because the DKDP is used in the form of a single non-scattering crystal, and above the Curie temperature.

In this article we shall review the physical principles of Phototitus, with emphasis on the special properties of the photoconductor. The properties of DKDP and the Pockels effect will only be mentioned briefly, as these have been treated before in articles in this journal on the TITUS tube [2]. We shall discuss the 'image-transfer characteristic', relating it to the photoconductor properties, and the spatial resolution. The article ends with a short review of possible applications, and a short comparison with related devices.

The Pockels effect of the DKDP plate

In a zero field a single crystal of DKDP is optically uniaxial, the optical axis z coinciding with the crystallographic axis c. It is therefore birefringent for light propagating in the a,b-plane, but light propagating in the c-direction is unaffected in polarization. An electric field E along the c-axis, however, induces a birefringence for light propagating in this direction proportional to the polarization $\varepsilon_c E$ of the crystal. This is the Pockels effect. The optical axes x,y in the a,b-plane bisect the axes a and b (*fig. 2*). In TITUS and Phototitus the c-axis is perpendicular to the plane of the plate. The light propagates in this direction, and is polarized along the a-axis by the polarizer P_1 (see fig. 1). After passing through the crystal (twice), the two components polarized parallel to x and y have acquired a phase difference ϕ proportional to $\varepsilon_c E l_C$, where l_C is the thickness of the crystal. The light thus emerges elliptically polarized. The amplitude of the light transmitted by the crossed analyser P_2 is proportional to $\sin \frac{1}{2}\phi$. For light polarized by P_1, the transmittance T of the system of DKDP plate plus polarizer P_2 is given by

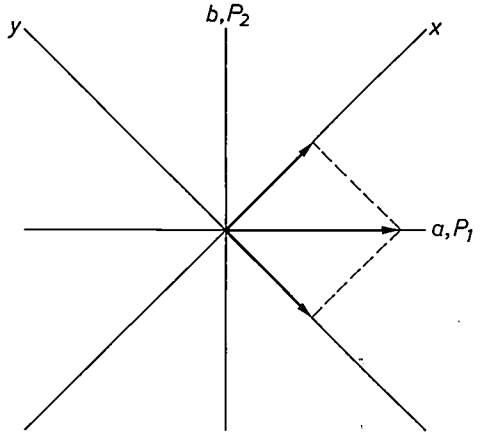


Fig. 2. Pockels effect in DKDP. An electric field E parallel to the c-axis induces birefringence in this direction, with optical axes x,y bisecting the axes a,b. Light polarized parallel to a by P_1 (see fig. 1) emerges elliptically polarized, as the components parallel to x and y have acquired a phase difference ϕ (proportional to $\varepsilon_c E$). The light thus acquires a component polarized parallel to b. This component is selected by P_2 (fig. 1).

$$T = \sin^2 \tfrac{1}{2}\phi = \sin^2 KV, \qquad (1)$$

where $V = El_C$ is the voltage between the faces of the DKDP plate. The factor K is proportional to ε_c, and independent of the thickness of the crystal [3].

[1] J. Donjon, F. Dumont, M. Grenot, J. P. Hazan, G. Marie and J. Pergrale, A Pockels-effect light valve: Phototitus. Applications to optical image processing, IEEE Trans. ED-20, 1037-1042, 1973 (No. 11). See also:
G. Marie and J. Donjon, Single-crystal ferroelectrics and their application in light-valve display devices, Proc. IEEE 61, 942-958, 1973 (No. 7).
G. Marie, J. Donjon and J. P. Hazan, in: Advances in Image Pick-up and Display 1, 226, 1974.
F. Dumont and D. Rossier, Le convertisseur optique Phototitus, and J. P. Hazan, Traitement des images, to be published in 1975 in Acta Electronica.

[2] G. Marie, Philips Res. Repts. 22, 110, 1967.
G. Marie, Philips tech. Rev. 30, 292, 1969.
'TITUS' is an acronym for 'Tube Image à Transparence Variable Spatio-temporelle'.

[3] More details are given in the article by G. Marie and J. Donjon, note [1].

Design and operation of Phototitus

In the experimental models of Phototitus that have been made in our laboratory, the active area of the multilayer structure is about 30×40 mm^2. The DKDP crystal is about 250 μm, the selenium layer about 10 μm in thickness. The dielectric mirror (M in fig. 1), the selenium layer (L) and the semitransparent electrodes (A_1 and A_2) are vacuum deposited upon the DKDP crystal. As in TITUS, the DKDP plate is attached by adhesive to a disc of calcium fluoride, to minimize the interaction between neighbouring points due to the piezoelectricity of the crystal. The entire unit is placed inside an evacuated container (*fig. 3*), where it can be cooled to about —50 °C by Peltier elements.

Fig. 4 shows an experimental arrangement for operating Phototitus. The polarizing beam splitter P, made of two calcite crystals, performs the functions of the polarizer P_1, the semitransparent mirror M' and the analyser P_2 in fig. 1.

In *fig. 5* the different stages of operation are represented by the potential across the device at some image point. The potential is given by curve a after a d.c. voltage, usually 150-200 V, has been applied between the electrodes A_1 and A_2. Most of the voltage drop appears across the selenium, because of the high value of ε_c in the DKDP. After writing-in curve b is obtained. The image is stored or read out with the electrodes short-circuited (curve c). For a sufficiently low level of the write-in light the voltage V_1 at the interface in stage

Fig. 3. Experimental model of Phototitus. The multilayer structure is mounted in an evacuated container; the selenium layer can be seen through the window. Two Peltier elements are also mounted in the container; these cool the system to —50 °C.

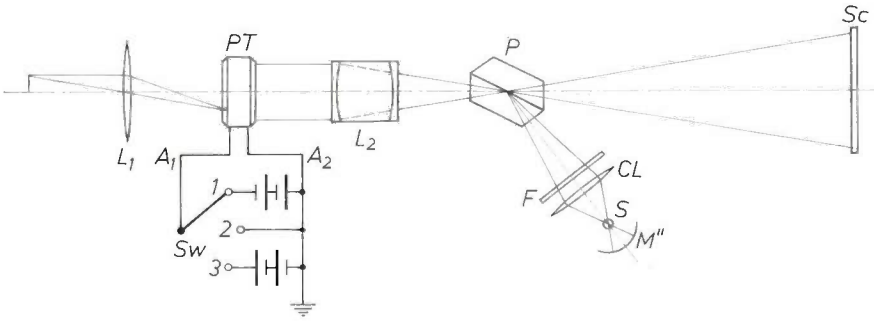


Fig. 4. Diagram of experimental arrangement for Phototitus. *PT* Phototitus. Writing-in takes place to the left of *PT*, read-out to its right. The polarizing beam-splitter *P* is made of two calcite crystals. L_1, L_2 lenses. *F* colour filter. *CL* condenser lens. *S* source of read-out light. *M''* spherical mirror. *Sc* projection screen. *Sw* switch for reversing the voltage on *PT*, or short-circuiting it.

c is proportional to the exposure at the image point considered (the 'exposure' is the product of the light intensity and the writing time). Finally, to erase the latent image, the whole area of the selenium is flooded with light while the electrodes are short-circuited (curve d). A selected area alone can also be erased by flooding that particular area. In fig. 5 details such as the presence of a dielectric mirror have been neglected; we shall return to these later on.

The photoconducting layer

Photoconducting amorphous selenium combines a number of properties that make it remarkably suitable for Phototitus. In the first place, vacuum deposition of a selenium layer is compatible with DKDP. This is not the case with CdS-type photoconductors, because of the high deposition or annealing temperature required to obtain reasonable sensitivity and low dark current. (CdS-type, i.e. II-VI compound, photoconductors are

used in several of the related devices, because of their high sensitivity in the visible range.) Secondly, as has been found in xerography, charge storage in amorphous selenium is very good, much better than in CdS-type photoconductors. In certain devices this may not be important, particularly when the images are not stored in the photoconductor (see p. 284). Thirdly, at the temperature of operation, —50 °C, the photoconductor is ambipolar; the $\mu\tau$ products for both electrons and holes are approximately equal and their value is high (μ is the mobility and τ the lifetime). This makes the device electrically symmetrical, and allows the reversal and subtraction of latent images. Finally, selenium has a sharp threshold for photoconduction in the short-wavelength part of the visible spectrum. This allows the input and output to be optically well separated.

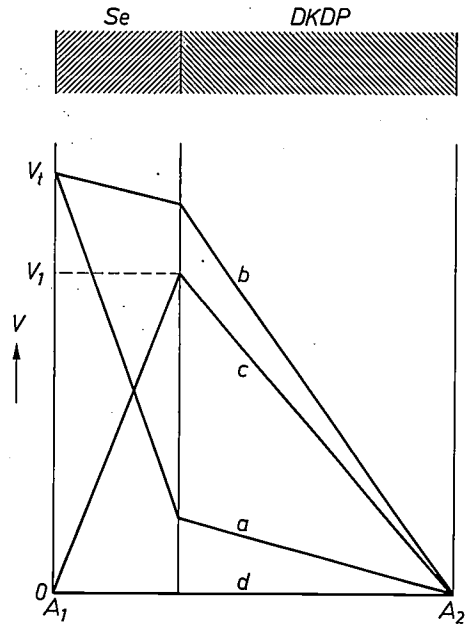


Fig. 5. Potential variation in the selenium and the DKDP, at some image point of the structure, for different stages of operation. After a voltage V_t has been applied (*a*), the optical image is written in (*b*). The electrodes are then short-circuited (*c*). In this stage the latent image is stored or read out. It can be erased by flooding the write-in side with light, with A_1 and A_2 short-circuited (*d*).

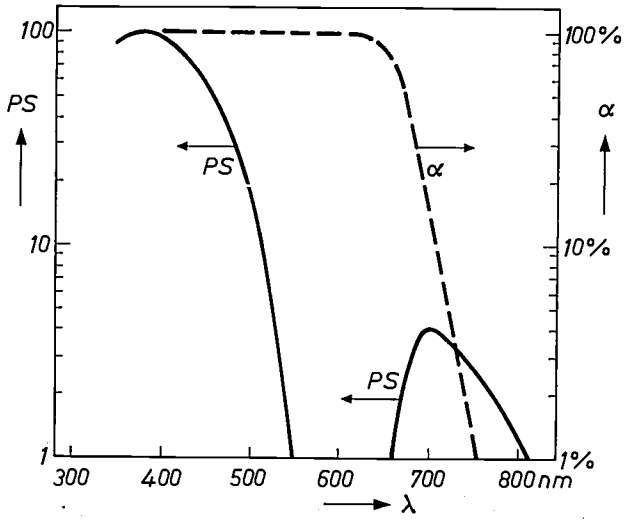


Fig. 6. Photosensitivity (*PS*) and optical absorption (α) of a 10-μm thick amorphous selenium layer coated by a gold electrode, as a function of the wavelength λ of the incident light. The left-hand part of the photosensitivity curve shows an intrinsic property of the selenium, and holds for both polarities of the voltage across the layer. The bump at the right is only present for a positive polarity (A_1 positive), and is due to holes photoemitted from the electrode A_1 into the selenium. The *PS* scale on the left is in arbitrary units.

In the following, we shall go a little further into the properties of amorphous selenium, when examining successively the *photoelectric conversion* of an optical image into a charge pattern (taking place in a thin layer at the selenium surface) the *transfer* of the charge pattern towards the interface and its *storage* at the interface.

Photoelectric conversion

The generation of free charge carriers in selenium by incident photons is a rather complicated process. The yield of this process, the 'quantum efficiency' η, is not only sharply dependent upon the wavelength of the light but also on the applied field.

Let us first consider the dependence on the wavelength. *Fig. 6* shows the photoconductivity spectrum and the optical absorption spectrum of a 10-μm layer of amorphous selenium. These results were obtained with a small fixed voltage across the layer, and under conditions similar to the operating conditions for Phototitus; with the layer at a temperature of —50 °C, and covered with a semitransparent electrode that partly absorbed the photons. The photoconductivity curve has a sharp edge near 550 nm, the absorption curve has one near 700 nm. Yellow light is completely absorbed in the layer, but it is ineffective in producing free carriers. In fact, there are photon-absorption processes in selenium that do not produce free carriers. From the data used for fig. 6 it can be deduced that the photons that do produce carriers efficiently are all absorbed within a layer less than one micron thick behind the electrode. The drop in photoconductivity near 550 nm is just a drop in quantum efficiency [4].

As mentioned above, the presence of an edge in the photoconductivity curve is rather fortunate: it permits a high optical isolation between input and output. The write-in light can be partly or wholly in the blue. If yellow light (wavelengths larger than 550 nm) is used for reading out, the latent image is little affected by it. This point will be further discussed later.

In selenium the quantum efficiency is rather low at —50 °C, as long as the applied field is weak. In fields higher than 10^4 V/cm, however, η increases strongly

[4] J. L. Hartke and P. J. Regensburger, Phys. Rev. 139, A 970, 1965.

with field-strength. In Phototitus, at low light levels, the field-strength in the selenium layer is of the order of 10^5 V/cm (100 V across a 10-μm layer). As we shall see below, our results indicate an efficiency of about 20% at such a field-strength. Of course, when the field in the layer decreases as charge builds up at the interface (see fig. 5, curve *b*), η decreases again. The field dependence of η is thus an important factor in the saturation behaviour of Phototitus. In our discussion of saturation, which we shall return to later, we have assumed that η depends on field-strength E in accordance with the relation:

$$\eta = \eta_0 \exp(-E_0/k\Theta + \beta E^{\frac{1}{2}}/k\Theta), \qquad (2)$$

proposed by D. M. Pai and S. W. Ing [5] and by M. D. Tabak and P. J. Warter [6]. In eq. (2), Θ is the absolute temperature, k Boltzmann's constant, and E_0 an activation energy at zero field. The factors η_0 and β, and also E_0, may depend on the wavelength λ of the light.

In the theoretical model on which eq. (2) is based, the probability that photons will generate a pair of free carriers at zero temperature and zero field is zero; η_0 is the probability of creating a pair of one hole and one electron, bound together by Coulomb forces. The hole and the electron may separate through thermal excitation, and this process is assisted by the electric field. The situation closely resembles that of *fig. 7*, representing the situation for an electron bound to a *fixed* positive point charge. The activation energy required to liberate the electron decreases with increasing field-strength. This is the 'Poole-Frenkel effect'. Elaboration of this model [5] [6] leads easily to eq. (2).

Field-induced transfer of the charge pattern

The efficiency of the transfer of the charge pattern across the bulk of the selenium layer is given by the ratio R of the lifetime τ to the transit time τ_{tr} of the charge carriers. If μ is the drift mobility of the carriers, l_L the thickness of the layer and V the voltage across it, the drift velocity is $\mu V/l_L$. Thus $\tau_{tr} = l_L^2/\mu V$, and $R = \mu\tau V/l_L^2$. Therefore, as far as the material is concerned, it is the product $\mu\tau$ that counts.

In selenium μ is the product of a microscopic mobility and a thermal activation Boltzmann factor associated with trapping by shallow centres [6] [7]. Thus μ is smaller at −50 °C than it is at room temperature. From the mobility data [8] for −50 °C, we find, with $V = 100$ V, $l_L = 10$ μm, a transit time of about 0.5 μs for holes and 100 μs for electrons. In practice, high-quality images can be formed by using a 10-μs write-in flash.

The lifetime τ is determined by trapping by deep centres. When a charge carrier is trapped by a deep centre, this becomes a recombination centre for carriers of the opposite sign.

The $\mu\tau$ product at room temperature, as derived from the values of μ and τ that have been found [9], is about 3×10^{-7} cm²/V for electrons, and ranges from about 3×10^{-7} to 10^{-6} cm²/V for holes. From the image-transfer characteristics of Phototitus, to be discussed later, we find values for $\mu\tau$ of 3 to 10×10^{-7} cm²/V at −50 °C, for both holes and electrons. To indicate the practical meaning of these values, we note that, with $V = 100$ V, $l_L = 10$ μm, they correspond to lifetime-to-transit-time ratios R between 30 and 100.

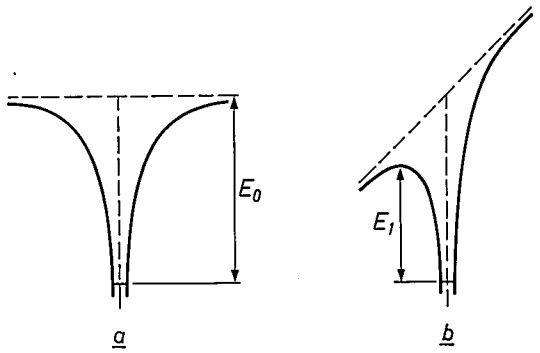


Fig. 7. Potential energy of an electron bound to a fixed positive charge, (*a*) without and (*b*) with an applied external field. The energy required to free the electron from the lowest bound state (E_0 in *a*, E_1 in *b*) is lowered when a field is applied. This is the Poole-Frenkel effect.

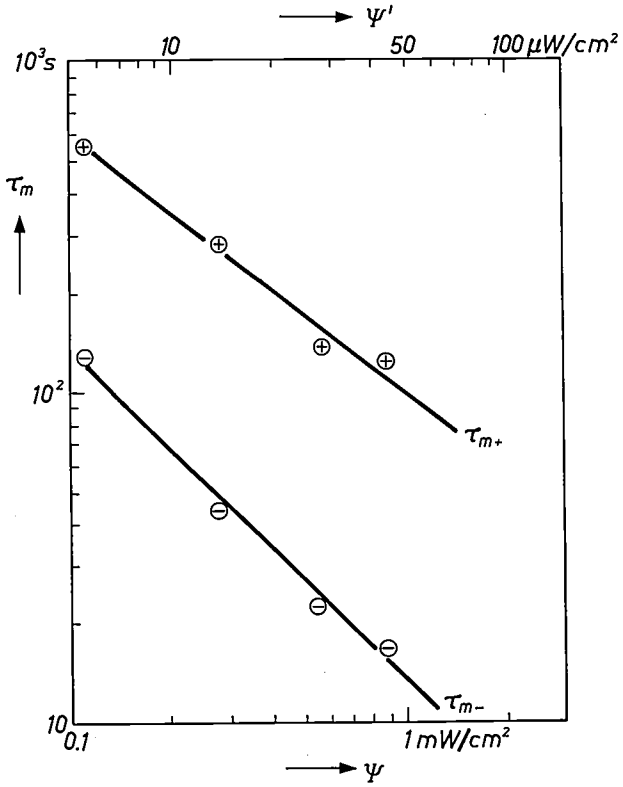


Fig. 8. Storage time τ_m as a function of the intensity ψ of the read-out light. The upper curve (τ_m+) is for a positive latent image, the lower curve (τ_m-) for the negative one. ψ' is the residual intensity of the light that has passed through the dielectric mirror and is absorbed by the selenium. Wavelength of the read-out light 579 nm.

Image storage

The persistence of a charge pattern built up near the interface between selenium and the dielectric mirror is limited by two processes. In the first place, charges may be thermally activated from the surface traps and diffuse laterally, which leads to progressive erasure and loss of definition. Secondly trapped charges may be annihilated by capture and recombination of free charges of opposite sign, leading to decrease in contrast without degradation of the definition. This is the dominant process in practice.

In the dark (without read-out light) thermal detrapping is very slow. This is a subsidiary advantage of operation at a low temperature. The storage time is of the order of an hour. On the other hand, when the DKDP is illuminated by the read-out beam, the residual transmission of the dielectric mirror ($> 5\%$ near the reflection maximum) allows some light to reach the selenium layer. Although the efficiency is low at the usual wavelength for observations ($\lambda \geqslant 560$ nm), this light generates free carriers, some of which recombine with charges of the latent image, thus erasing it (the second process mentioned above). *Fig. 8* shows the storage time τ_m as a function of the intensity ψ of the read-out light, with logarithmic scales. The curve for a positive latent image (the upper curve) has a slope less than unity. This can be explained by assuming that some of the holes generated by the read-out light disappear by a process of bimolecular recombination; if this were the only process, the carrier density would be proportional to the square root of the luminous intensity [10] and the slope would be 0.5. For a negative latent image erasing is much faster (lower curve). This can be explained by an additional recombination process, due to holes injected into the selenium by photoemission from the electrode A_1 [11]. In the case of a gold electrode, only holes are photoemitted. The photosensitivity spectrum of this process is represented by the bump in the red region in fig. 6. The photoinjected holes are attracted by the negative charge of the latent image. Their density is proportional to the read-out light intensity, leading to a mean storage time inversely proportional to this intensity. In both cases there is no sign of loss of definition during the erasure.

Image-transfer characteristic

It is important in practice to know the transmittance T of the system for the read-out light as a function of the exposure H on the write-in side. This function enables the exposure necessary to obtain a certain contrast to be determined. Such an 'image-transfer characteristic' for exposure wavelengths centred around 401 nm and an applied voltage of 150 V (A_1 positive), is shown in *fig. 9*. The characteristic is rather similar for the opposite polarity, which demonstrates the ambipolar nature of the selenium layer at the temperature of operation.

The experimental characteristics are generally S-shaped, and three zones can be distinguished: a residual-transmittance zone, an intermediate zone and a saturation zone. When the exposure approaches zero, the transmittance tends to a constant level, owing to stray light scattered or reflected by elements of the optical system. The curve in fig. 9 was obtained with a very low level of stray light, and does not show this effect. When this level is not so low, the same curve is obtained after the residual transmittance is subtracted from the experimental T-values. After correction for stray light, T is proportional to the square of the exposure, for small exposures (intermediate zone). This corresponds to the first-order expansion of eq. (1), where V, the voltage across the DKDP, is proportional to the exposure for small exposures. For large exposures the transmittance saturates to a constant value

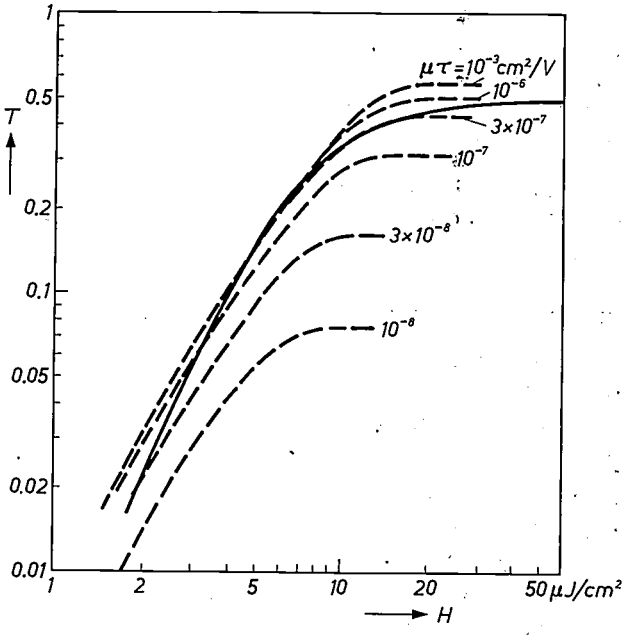


Fig. 9. Image-transfer characteristic. Transmittance T for the read-out light is plotted against the write-in exposure H. Write-in wavelength 401 nm. Applied voltage 150 V. Solid curve: experimental results; dashed curves: theoretical predictions for different values of $\mu\tau$ and a value of 0.20 for η in a field of 10^5 V/cm.

[5] D. M. Pai and S. W. Ing Jr., Phys. Rev. **173**, 729, 1968.
[6] M. D. Tabak and P. J. Warter Jr., Phys. Rev. **173**, 899, 1968.
[7] W. E. Spear, Proc. Phys. Soc. B **70**, 669, 1957.
[8] J. L. Hartke, Phys. Rev. **125**, 1177, 1962.
J. Schottmiller, M. Tabak, G. Lucovsky and A. Ward, J. non-cryst. Solids **4**, 80, 1970.
See also the article by W. E. Spear, note [7].
[9] See the articles of note [8], and also M. D. Tabak and W. J. Hillegas, J. Vac. Sci. Technol. **9**, 387, 1971.
[10] See for instance R. H. Bube, Photoconductivity of solids, Wiley, New York 1960.
[11] J. Mort, F. W. Schmidlin and A. I. Lakatos, J. appl. Phys. **42**, 5761, 1971.

T_s, depending upon the applied voltage. The presence of stray light together with the saturation of the transmittance limit the maximum contrast (ratio of transmittances) obtainable at a given voltage.

Stray light can be reduced to a level lower than 1% of the read-out intensity by careful design and adjustment of the optical equipment and by depositing an antireflection coating on the optical components of the device. Furthermore, when a very low residual transmittance is required, it is necessary to use a read-out beam accurately perpendicular to the DKDP single crystal, in order to avoid unwanted depolarization by the natural birefringence of the material [12]. With this precaution, we have obtained a maximum contrast much higher than 100 to 1.

Saturation occurs, first of all, because the potential drop across the selenium (see fig. 5, curve *b*) eventually disappears as charge accumulates at the interface, so that no more carriers will drift across the layer. If that were all, the relation between T_s and V_t would be given directly by eq. (1). The matter is however complicated by several factors. First, the dielectric mirror reduces the voltage across the DKDP to 2/3 of the voltage at the mirror-selenium interface (see *fig. 10*). Secondly, charge carriers caught in deep traps in the bulk of the selenium give a space charge, and thus a curvature of the potential characteristic (also indicated in fig. 10). Finally, in accordance with eq. (2), the quantum efficiency η falls off rapidly as saturation is approached.

We have worked out a theoretical model that takes these factors into account, and we have calculated *T-H* curves and T_s-V_t curves from it. The main parameters in the calculation are $\mu\tau$ and the value of η at a particular field-strength. Fig. 9 shows, besides the experimental curve, some calculated *T-H* curves. A change in the value of η is equivalent to a change in exposure and therefore shifts the curves in the horizontal direction. Fitting the experimental and calculated curves in this way, for a wavelength of 401 nm and a temperature of −50 °C, yields a value of 0.2 ± 0.03 for η at a field-strength of 10^5 V/cm. With the same conditions of field and temperature Pai and Ing found a value of 0.3 [5].

The position of the saturation plateau in fig. 9 yields directly an estimate for $\mu\tau$ of about 10^{-6} cm²/V. A more accurate way of determining $\mu\tau$ consists in fitting the calculated T_s-V_t curve to the experimental data, as in *fig. 11*. The best value thus found is 3×10^{-7} cm²/V.

The set of calculated curves in fig. 11 is useful for evaluating the minimum value of $\mu\tau$ compatible with the contrast required for a given application. This problem is of practical importance in the technology of the converter, since for long-term reliability of the device the selenium layer has to be stabilized by doping it with glass-forming additives. However, as $\mu\tau$ is usually affected by even a small quantity of impurities [13], a compromise has to be found which depends on whether high contrast or long life is most important.

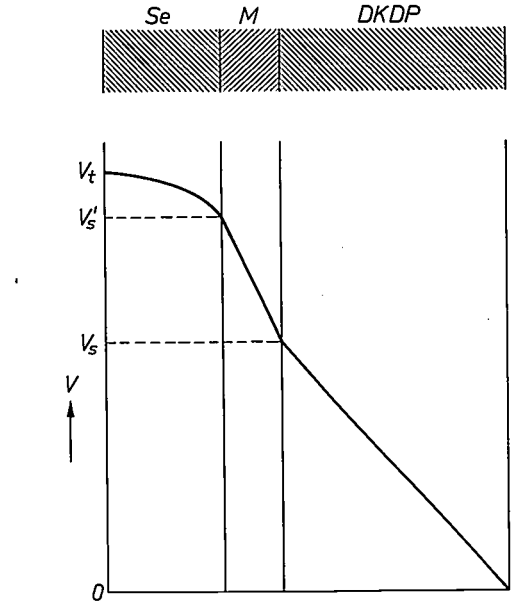


Fig. 10. The saturated transmittance T_s is determined via eq. (1) by the saturation value V_s of the voltage across the DKDP crystal. V_s is lower than the tube voltage V_t because: *a*) V_s is only $\frac{2}{3} V_s'$, where V_s' is the voltage at the selenium/mirror interface; *b*) trapped charge carriers give a space charge, bending the potential curve in the selenium; *c*) η depends strongly on the field. (This last effect is not taken into account in the figure.)

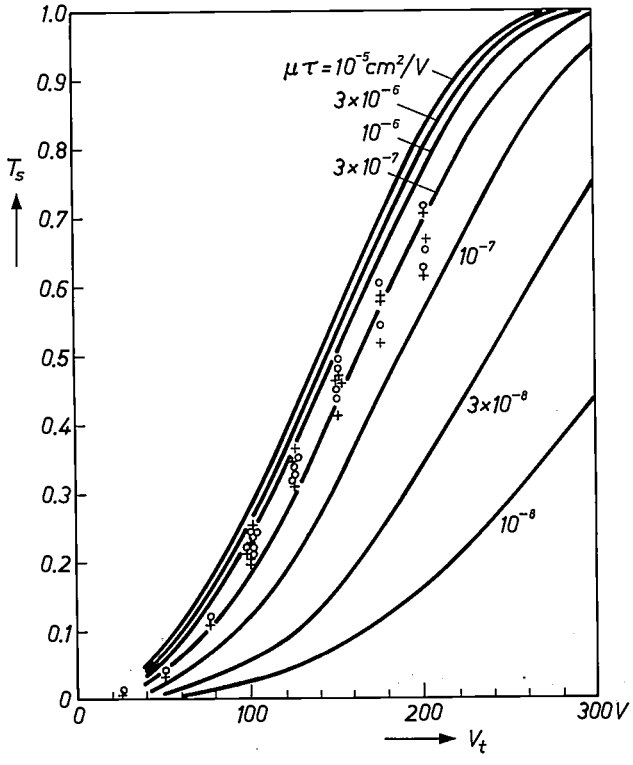


Fig. 11. Saturation transmittance T_s plotted against tube voltage V_t. Experimental points: + electrode A_1 positive, ○ electrode A_1 negative. The solid lines are theoretical curves for different values of $\mu\tau$.

Resolution

In many applications high definition of the stored image is a crucial requirement. Calculations made by J. Donjon of this laboratory have shown that the theoretical resolution is 40 line pairs per mm with a modulation of 10%, or 80 l.p./mm with a modulation of 5%. As in related devices, the main source of point spreading is the fringing of the stored electric fields [14]. By using incoherent read-out light at 590 nm, we have achieved a resolution of more than 75 l.p./mm on Phototitus (see *fig. 12*). This value corresponds to 5700 elements per horizontal line, since the DKDP crystals have a width of 38 mm. This kind of high-quality image can be stored and displayed during several tens of seconds without degradation of the resolution with a total incident luminous flux of about 1 lumen on the DKDP. This enables an output illuminance of about 60 lux on the surface of a 9×11 cm screen corresponding to standard photographic film. It has not previously been possible to store an image with such a high resolution at higher luminous flux (e.g. more than 10 lumens) for a long time, because of the existence of the various recombination processes in the storage layer discussed above. We have however managed to store and display high-illuminance images of 100 lux on a $\frac{1}{2}$ m² screen for 30 seconds by depositing an 'optical shield' between the dielectric mirror and the photoconducting layer. But in this kind of arrangement the resolution was never better than about 10 l.p./mm because of the finite — although very small — lateral photoconductivity of the optical shield layer.

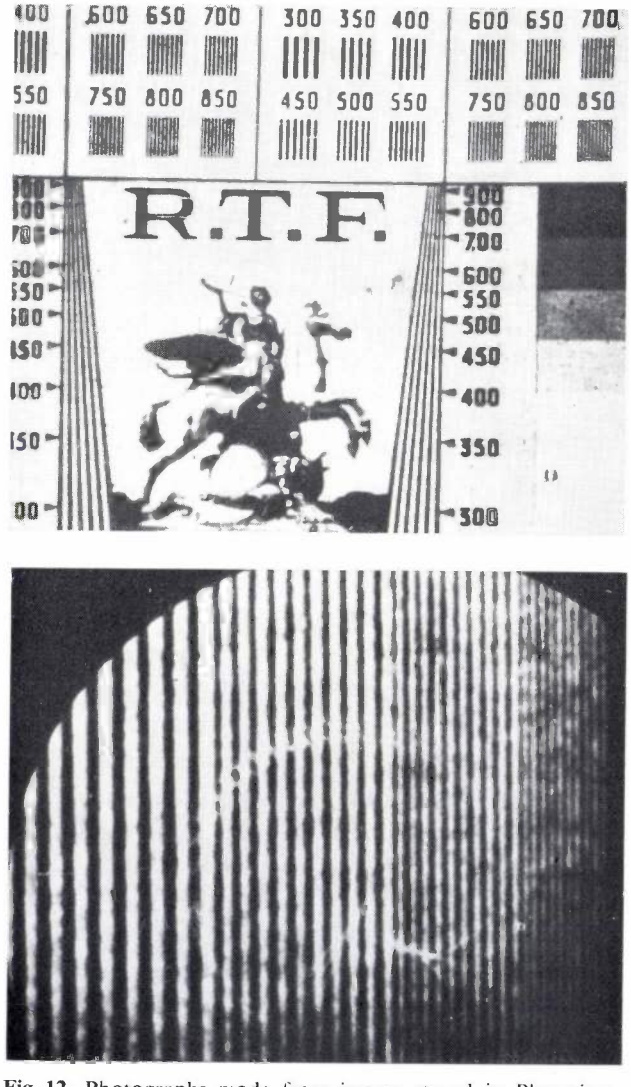


Fig. 12. Photographs made from images stored in Phototitus. *Above:* standard television test pattern. *Below:* high-definition test pattern. The pitches of the vertical lines are 42 μm, 35 μm, 24 μm and 16 μm, from left to right in the lower photograph. The 16 μm pitch is still clearly discernible: the limiting resolution attained in practice is 75 l.p./mm.

Applications

Some examples of image processing by Phototitus are shown in *figs. 13*, *14*, *15* and *16*. In these examples the read-out was performed with ordinary incoherent light.

Fig. 13*b* shows the result of subtracting a uniform charge distribution from the latent image of fig. 13*a*. This is done by simply making a uniform exposure with reversed voltage. Contours of a particular grey level in the original show up as black contours in the treated picture. A similar result can be obtained by subtracting an 'optical bias' (i.e. a phase shift) by means of a Bravais compensator or a rotable quarter-wave plate in series with the device. Fig. 13*c* and *d* indicate how charge variations along a line in the latent image are translated into variations of transmittance. In fig. 13*c* the original variation is reproduced only slightly distorted. In fig. 13*d*, however, where a 'uniform image' has been subtracted, lines of zero charge (lines of a particular grey in the original) become black lines.

It will be clear from fig. 13*d* that a complete 'positive-to-negative image conversion' can also be obtained, by simply subtracting more charge than in the case of fig. 13*d*.

Writing a picture into Phototitus and subtracting it again after a slight shift in a particular direction amounts to determining the spatial derivative of the greyness in that direction. Differentiation can be used to enhance contours or small details in pictures (fig. 14*a* and *b*); in such cases the subtraction is incomplete to preserve part of the original image. As indicated in fig. 14*c*, positive and negative derivatives will all show up as white lines because of the quadratic nature of the photographic plates. However, as is seen in fig. 14*d*,

[12] See the article by G. Marie and J. Donjon, note [1].
[13] See the article by J. Schottmiller *et al.*, note [8].
[14] D. S. Oliver and W. R. Buchan, IEEE Trans. **ED-18**, 769, 1971.

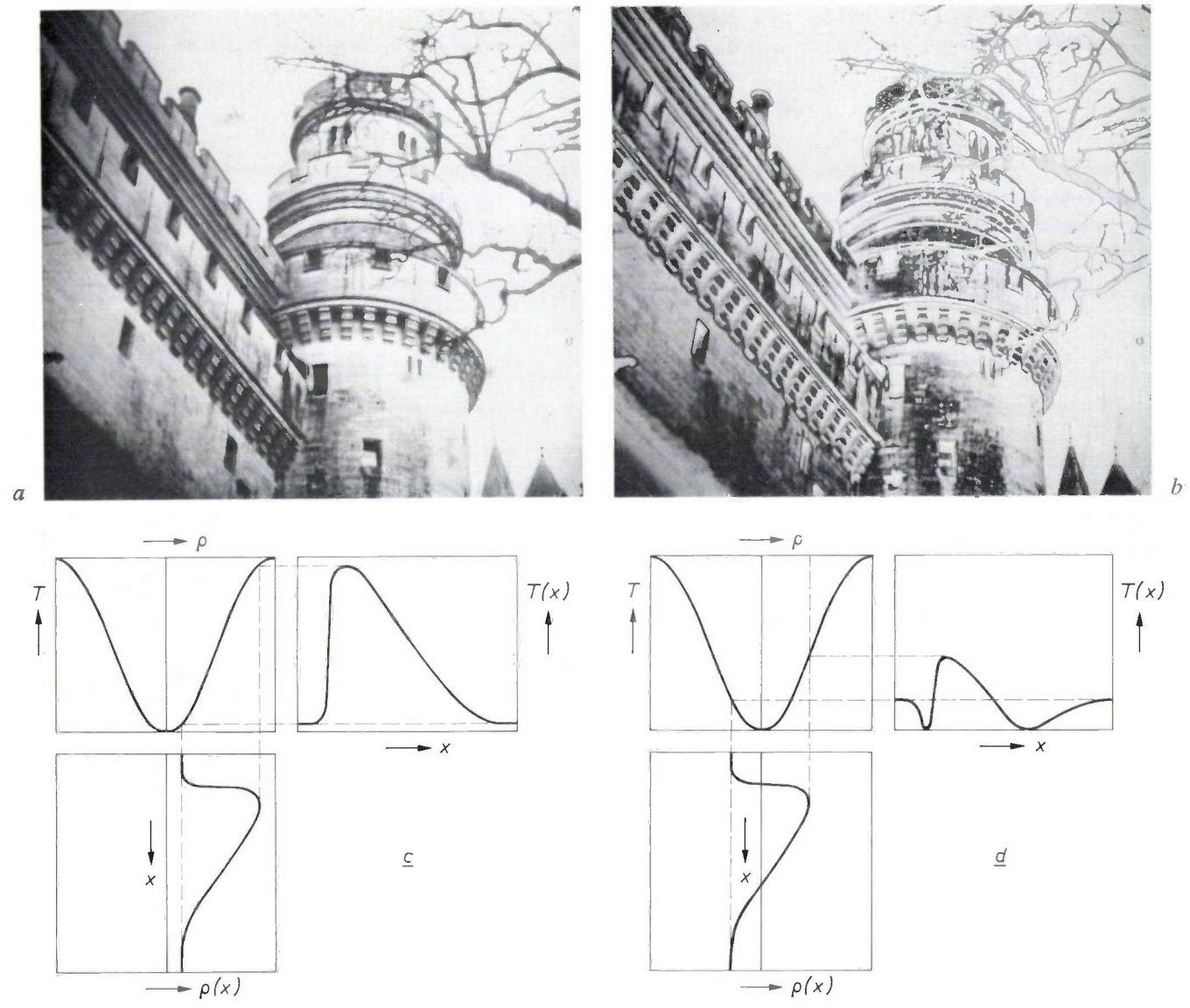


Fig. 13. Obtaining lines of a particular grey level. *a*) Original picture, *b*) picture obtained after subtracting a 'uniform image' (i.e. a uniform exposure). The diagrams show how a charge variation $\varrho(x)$ in the latent image is translated into a transmittance $T(x)$, for the original (*c*) and for the treated image (*d*).

negative derivatives can be made to show up as black and the positive ones in white, by adding a uniform image to the differentiated image.

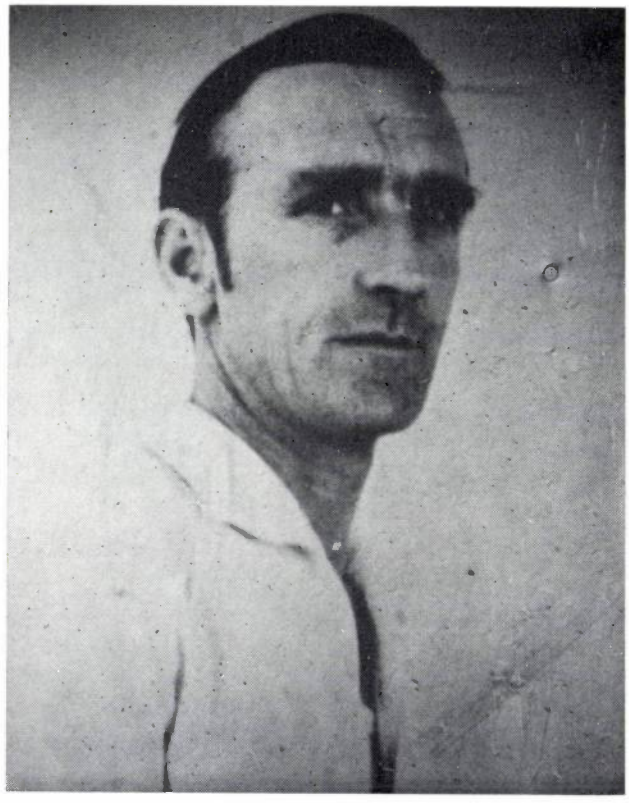
Fig. 15 shows the effect of subtracting a slightly blurred picture from a sharp original. The result is a contour enhancement in all directions. This procedure amounts to an approximate determination of the second derivative (fig. 15*c*), or, more precisely, the Laplacian $\partial^2/\partial x^2 + \partial^2/\partial y^2$ of the greyness (where x and y are the coordinates in the plane of the picture).

These almost instantaneous methods of determining constant grey-level contours and of contour enhancement may be useful where pictures have to be scrutinized for particular details, as in medical X-ray diagnosis.

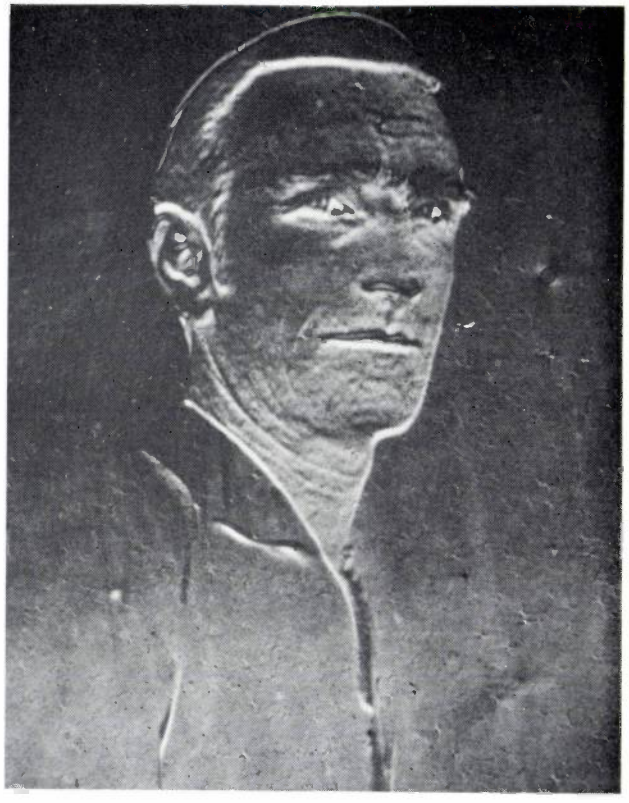
In cases where a scene is almost but not completely stationary, changes in the situation may be detected by subtracting two successive pictures. In the limit this amounts to determining the time derivative. This feature could be applied in security monitoring, or in the analysis of pictures from weather or Earth-resources satellites.

The addition of pictures is the basis of integration and averaging. For instance, restoration of photographs or television images affected by optical or electrical noise during the pick-up or transmission stages can be made in this way. This process is illustrated in fig. 16.

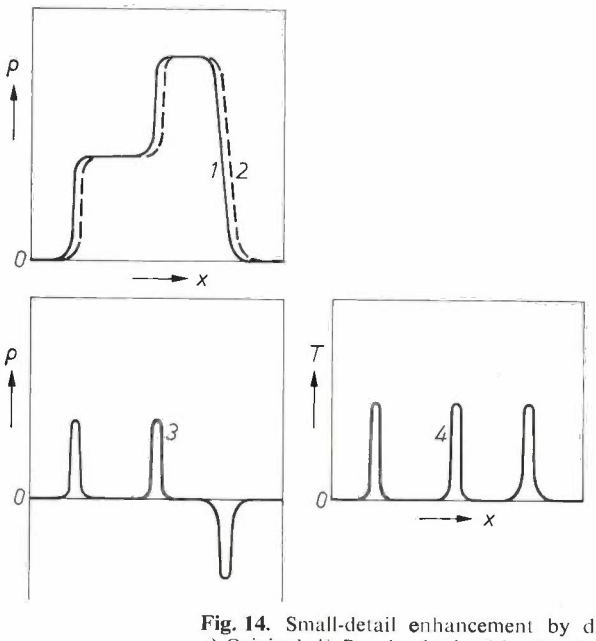
The most important application of Phototitus will probably be in image processing with coherent light. Since the advent of the laser many schemes for coherent optical processing of images and data have been



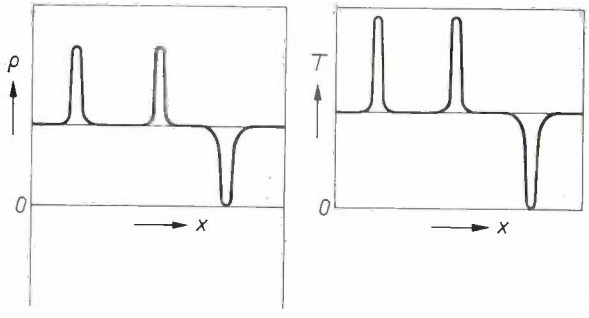
a



b



c



d

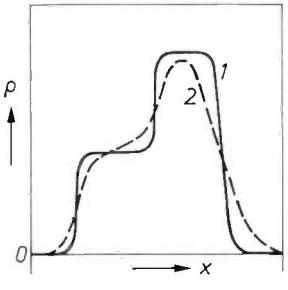
Fig. 14. Small-detail enhancement by differentiation with part of the original preserved. *a*) Original. *b*) Result obtained by (partial) subtraction, from the original, of the same picture. *c*) The principle of first-order differentiation. The charge distributions of the original (*1*), the shifted (*2*) and difference (*3*) images are indicated schematically. Both leading and trailing edges show up as white because the photographic film is a quadratic detector. *d*) Adding a uniform image will make one kind of edge show up as white, the other black.

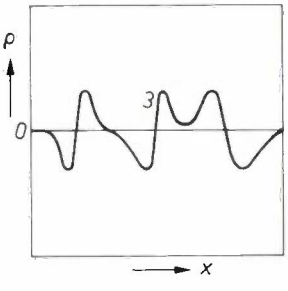
devised [15]. In seismology, for instance, the spatial frequencies present in a particular pattern are of interest and these can be identified by the method indicated in *fig. 17*. The Fourier transform of the transparency T will appear in the focal plane F of the lens L if the beam B is coherent and T is in the other focal plane of L. Thus gridlike patterns in T can be identified by spots in F. Another of the many applications of coherent optical processing is in character recognition, as indicated

[15] See for instance A. Vander Lugt, Optica Acta **15**, 1, 1968, and also J. W. Goodman, Introduction to Fourier optics, McGraw-Hill, New York 1968.

a

b





c

Fig. 15. Omnidirectional contour enhancement. *a*) Original picture and *b*) treated picture, obtained by subtracting a blurred picture from the original. *c*) Charge distributions of the original (*1*), the blurred picture (*2*), and the difference (*3*); the curve *3* is approximately the second derivative of *1*.

in *fig. 18*. When a holographic filter of a character is prepared as in fig. 18*a*, this character if present in some text on a transparency *T*, will be identified in the arrangement of fig. 18*b* by a spot on the screen *S*. This method can be extended for fast automatic reading [16].

Until now there has been one great difficulty with such methods. Because of the speckle effect observed on laser illumination of a stationary scattering object, coherent optical processing of images on data displayed on a scattering medium (e.g. paper) is not in general possible. In practice high-quality photographic film is used. This, however, excludes such applications as fast reading of ordinary print on paper, which would be very useful for feeding printed information into a computer. Making a film of such a print, to be read perhaps only once, is a slow and wasteful process. In short, the method is inhibited by lack of 'real-time, re-usable film'.

It will be clear from the foregoing that Phototitus offers a solution to this problem. Print projected on the write-in side can be read out by coherent light. *Fig. 19* shows the result of an early character-recognition experiment in which Phototitus was used in this way. We estimate that more than 20 000 characters per second could be read with a developed version of the method. This is about ten times faster than the fastest optical reader generally available.

Finally, we would like to mention the possibility of using Phototitus to eliminate the 'zero-order diffraction term'. In many coherent optical processing experiments, e.g. those of figs. 17 and 18, the zero-order term of the optical Fourier-transform, representing the average picture amplitude, carries no interesting information: it yields a much brighter spot than the useful information, and scattering by the filter plate of a small fraction of this spot could be a serious source of noise. With Phototitus, the zero-order term can be eliminated by subtracting from the latent image a 'uniform' image of amplitude equal to the average amplitude of the original, or by using an optical bias. A reduction of the zero-order diffraction term by several hundred times has been achieved.

Comparison with related optical devices

In this final section we shall briefly compare Phototitus with other devices that have been developed in recent years for image processing, which are also based on the idea of combining photoconductivity with an electro-optical property.

In *Table I* we have collected the most important data for different types of converters, including Phototitus. Comparison of these data should be made with some reserve, because the conditions of measurement were

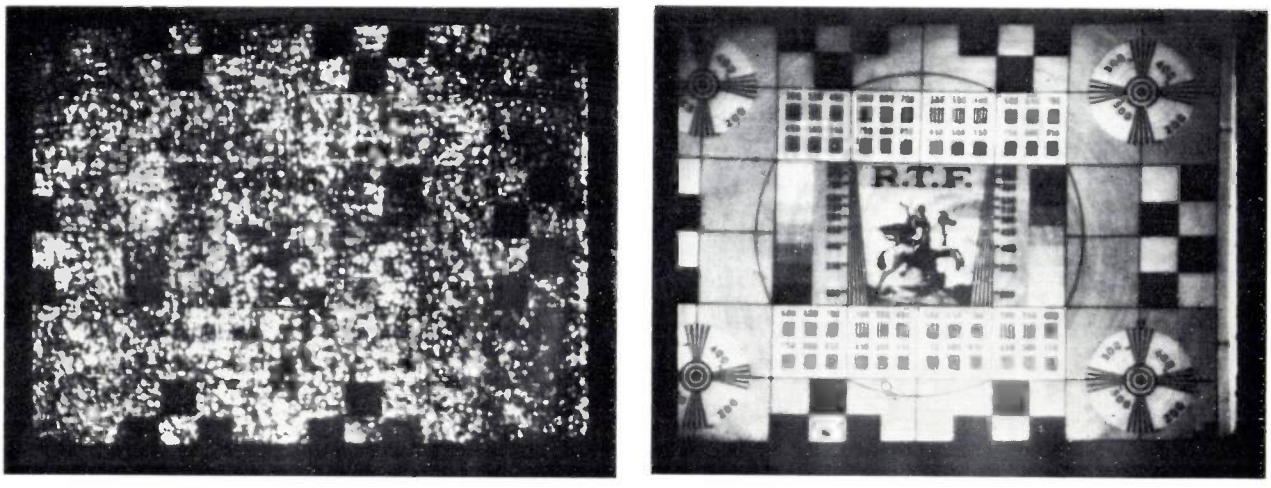


Fig. 16. Signal-to-noise enhancement by addition, on Phototitus, of images affected by uncorrelated noise. *Left:* image obtained by placing a diffusor in front of a picture and projecting it on Phototitus. *Right:* image obtained with the diffusor rotating during the exposure. The rotating diffusor simulates uncorrelated noise, which is added during the exposure in the right-hand case. The total exposure is the same for both cases.

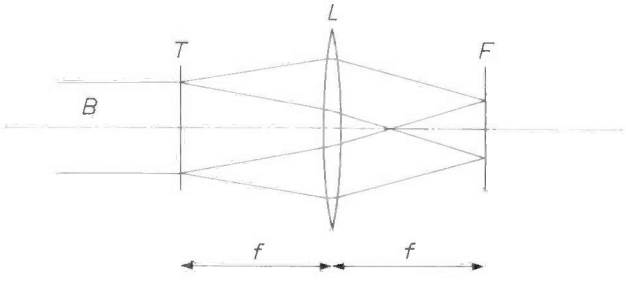


Fig. 17. Analyses of spatial frequencies by coherent optical processing. *T* subject, in the form of a transparency, to be analysed. *B* coherent beam of light. *L* lens. If *T* and *F* are in the front focal plane of *L*, the two-dimensional spatial Fourier transform of *T* will appear in the back focal plane *F*. Grid-like structures in *T* give dots in *F*.

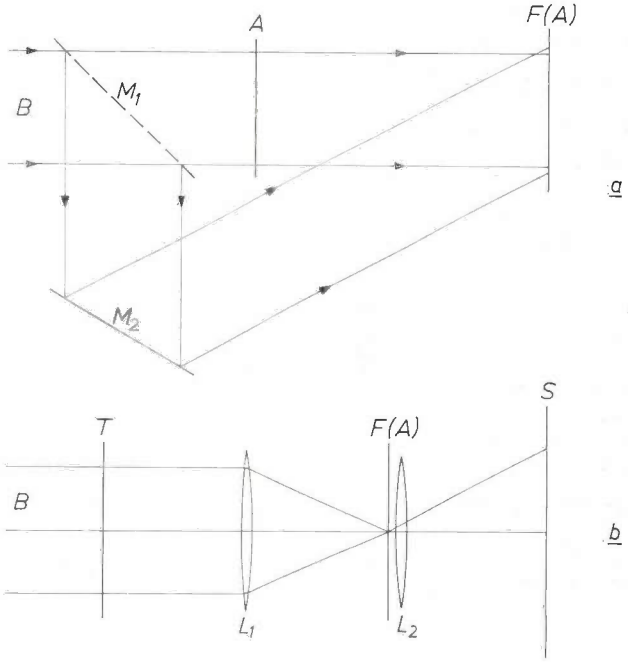


Fig. 8. Character recognition by autocorrelation. *a*) Preparation of a holographic filter $F(A)$ of the letter A, presented as a transparency. B coherent beam of light. M_1 and M_2 mirrors (M_1 semi-transparent). *b*) Searching through a text on the transparency T for the letter A. $F(A)$ holographic filter, as prepared in (*a*). L_1, L_2 lenses. S screen. Each letter A in T will give a separate autocorrelation spot in S, related to its position.

not always reported in detail, and also because of differences in the physical principles employed.

Devices 1 and 2 in Table I are of the 'Fe-Pc' type in which a photoconductor (Pc) is combined with a ferroelectric material (Fe). (By analogy Phototitus would be of the 'Pa-Pc' type.) In these devices images can be stored in the ferroelectric material because of the opposition of the coercive force to the relaxation of 'flipped' domains. They are therefore operated at a temperature below the Curie point. Phototitus, however, is operated above the Curie point, so there is no danger of coherent light being scattered by domain walls. Devices based on ceramics [17], which are most useful in other applications, do of course scatter coher-

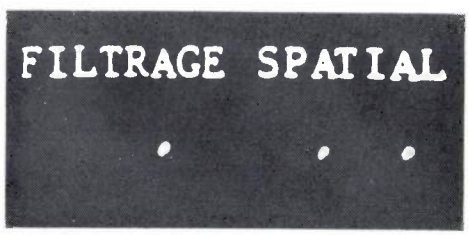


Fig. 19. Automatic reading of the letter A with the method of fig. 18, where, instead of A in fig. 18*a*, and T in fig. 18*b*, Phototitus was used, translating print on paper into a transparency. *Above:* text. *Below*: dots identifying the letter A.

[16] See for instance M. Treheux, in: Applications de l'holographie, Proc. int. Symp., Besançon 1970, paper No. 13.3.

[17] J. R. Maldonado and A. H. Meitzler, Proc. IEEE **59**, 368, 1971.
D. W. Chapman, J. Vac. Sci. Technol. **9**, 425, 1972.
W. D. Smith and C. E. Land, Appl. Phys. Letters **20**, 169, 1972.

ent light completely; for this reason they are not considered here.

In the devices 3 and 4, the 'Ruticon' and 'Fericon', the electro-optical effect is not a bulk effect, as in Phototitus and the other devices, but a surface effect. The operation of the Ruticon is based on the surface deformation of a membrane, whereas the operation of the Fericon is based on the deformation of a metallic layer deposited on the surface of a ferroelectric material. A disadvantage of using a surface electro-optical effect is that only a relatively low contrast can be obtained.

hundred volts, whereas the $Bi_{12}SiO_{20}$ PROM needs voltages of the order of a thousand volts. The maximum read-out transfer ratio of Phototitus is higher than in the PROMs, because of the high transparency of the DKDP single crystal and the electrode A_2 in fig. 1, and also because of the possibility of approaching a 90° phase retardation, on account of the high electro-optical sensitivity of DKDP and the high transfer efficiency of selenium.

The DKDP crystals used in Phototitus have an area five times that of the $Bi_{12}SiO_{20}$ crystal used in

Table I. Data for several types of optical converters. $V_{\lambda/2}$ applies only to devices using the Pockels effect (the PROMs and Phototitus). It is the voltage across the electro-optic crystal necessary to obtain a phase difference of π between the x- and y-components (see fig. 2), i.e a 100% transmittance; $V_{\lambda/2}$ equals $\pi/2K$, where K is the factor used in eq. (1).

The 'writing sensitivity' is the exposure (energy per cm^2) required on the write-in side to obtain a 'read-out transfer ratio' of 10 or 20% of its maximum value. 'Read-out transfer ratio' is the intensity of the read-out light leaving the device divided by the intensity of the incident read-out light. The term 'grating transfer ratio' is used when the only part of the read-out light counted is that present in a beam diffracted by a latent image in the form of a grating.

Type of device	Active area	$V_{\lambda/2}$	Sensitivity for writing	Write-in wavelength	Max. resolution	Max. contrast	Max. read-out transfer ratio	Storage time (dark)	Grey scale
1. Fe-Pc $Bi_4Ti_3O_{12}$ + ZnSe [18]			1000 μJ/cm²	514.5 nm	70 l.p./mm	5 : 1	0.01 % [*]	days	limited (binary)
2. Fe-Pc $Bi_4Ti_3O_{12}$ + PVK	0.16 cm²		2500 μJ/cm²		114 l.p./mm				limited (binary)
3. Metal-plated Ruticon + PVK [19]	> 20 cm²		30 μJ/cm²	632.8 nm	40 l.p./mm		1.8 % [**]		very good
4. Fericon [20]					57 l.p./mm	2.2 : 1			
5. PROM [14] ZnS or ZnSe	2 cm²	13 kV	10 μJ/cm²	340 nm	85 l.p./mm	10 : 1	2%	100 h	good
6. PROM [21] $Bi_{12}SiO_{20}$	2.25 cm²	3.9 kV	12 μJ/cm²	400 nm	80 l.p./mm	1000 : 1	8%	2 h	good
7. Phototitus	12 cm²	150 V	10 μJ/cm²	401 nm	70 l.p./mm	> 100 : 1	50%	½ h	very good

[*] Grating transfer ratio at 632.8 nm.
[**] First-order grating transfer ratio for a latent image of 20 l.p./mm.

[18] S. A. Keneman, G. W. Taylor, A. Miller and W. H. Fonger, Appl. Phys. Letters **17**, 173, 1970.
[19] N. K. Sheridon, IEEE Trans. **ED-19**, 1003, 1972.
[20] C. E. Land and W. D. Smith, Appl. Phys. Letters **23**, 57, 1973.
[21] P. Vohl, P. Nisenson and D. S. Oliver, IEEE Trans. **ED-20**, 1032, 1973.

Devices 5 and 6 are 'PROMs' (Pockels Read-out Optical Memories). In these the two functions are performed by one material, which is photoconductive as well as electro-optic.

Inspection of Table I shows that only the PROM based on a single crystal of $Bi_{12}SiO_{20}$ could be a serious competitor to Phototitus for incoherent-to-coherent optical conversion. Phototitus, however, has the advantage of a much higher electro-optical sensitivity. It can be operated by switching voltages of the order of a

PROMs; they can nevertheless be made sufficiently homogeneous to give good images over the whole area.

Finally, we should note the importance of the physical separation of the photoconductivity and the electro-optical effect, which permits both to be optimized. As an example, we have seen that, in the case of Phototitus, it is possible to select an ambipolar photoconductor, enabling us to manipulate both positive and negative charge images, and to subtract them accurately. This is not possible with the PROM based on

ZnS [14]. Image subtraction was not reported in the case of the $Bi_{12}SiO_{20}$ PROM [21].

Function separation is also essential for preserving the image while it is read out. As explained in the foregoing, read-out light can produce free carriers in the photoconductor. This unwanted effect becomes serious when the read-out beam passes through the photoconducting layer itself as in PROM, whereas in Phototitus it is reduced by two orders of magnitude by the dielectric mirror. As a result, Phototitus allows an image to be observed for a few minutes with yellow-orange light (590 nm), where the eye is very sensitive, whereas images in the $Bi_{12}SiO_{20}$ PROM are very rapidly erased if the read-out light is not red.

Summary. Phototitus, a device for image processing, consists of a 30 × 40 mm single-crystal slice of DKDP, 250 μm thick, covered with a dielectric mirror and a layer of amorphous selenium, 10 μm thick. Semitransparent electrodes cover both faces of the structure. With a voltage of 150-200 V between the electrodes, an optical image projected on to the selenium is converted into a latent charge image at the interface. The Pockels effect in the DKDP permits the latent image to be read out with polarized light. A large Pockels effect is obtained by cooling the DKDP to just above its Curie point (−50 °C) by Peltier elements. The storage time of the latent image in the dark is tens of minutes. Images can be added by writing them in one after the other. An image is subtracted by writing it in with the voltage reversed. This possibility is due to the ambipolarity of the selenium. Thus images can be integrated, averaged or differentiated with respect to position or time, leading to such applications as the elimination of uncorrelated noise, contour enhancement and the detection of changes in an image. An important application could be as 'real-time re-usable film', i.e. the instantaneous conversion of print into a 'transparency' that can be processed by coherent light, e.g. for character recognition. A brief comparison is made with other optical converters. The relation between the characteristics of the device and the intrinsic properties of the photoconductive selenium layer is discussed.

Philips tech. Rev. 34, 288-293, 1974, No. 10

The frequency-analog signal as a basis for measurement and control

D. Gossel

Previous articles in this journal have dealt with results obtained at Philips Forschungslaboratorium Hamburg in research on frequency-analog measurement and control systems. In these systems the information is contained in the frequency of the signals used, which considerably reduces interference in transmission. The systems can also easily be incorporated in large industrial installations, which nowadays generally use digital data processing. The present article, which can be considered as a synopsis of the features of such systems, was first published in German in the book 'Unsere Forschung in Deutschland II' (Our Research in Germany) where it appeared as the introduction to four publications on applications of the new principle [*].

Amplitude-analog or frequency-analog measurement?

Information about a physical process is usually obtained from measurements of the variables. Sometimes the indicating instrument and the sensor can be combined in a single unit at the site of the measurement, as in the mercury thermometer, for example. Often, however, especially in industrial processes, the measured quantity x first has to be converted *in situ* into a quantity y that can readily be transmitted. At the receiving end the signal y, which has generally been modified by interference during transmission, is converted into a quantity z that can be read off, recorded, stored in a memory or arithmetically processed.

Electrical and pneumatic quantities are particularly suitable for transmission; typical signals are current values, voltages and resistances, or air pressure. The observed indication (amplitude) can then follow the behaviour of the measured quantity x and thus provides a quantity analogous to the magnitude of x — hence the term 'analog' measurements. It will be useful here to use the more exact term 'amplitude-analog' to distinguish it from other conceivable types of analog.

In practice amplitude-analog measurements are found to be highly sensitive to interference, and a good deal of auxiliary equipment is necessary to convert the received signal into a form suitable for digital processing, nowadays an increasingly common requirement. These difficulties led to proposals that, instead of the amplitude, the frequency or period of an a.c. signal should be used as the signalling quantity. The frequency of an electrical signal is then very little affected by interference in transmission channels, and furthermore a simple physical relationship exists between many measured quantities x and their frequency or period.

Since 1964 investigations have been carried out at Philips Forschungslaboratorium Hamburg on measurement and control systems based on the use of such 'frequency-analog' signals [1]. A summary of the progress made up to 1969 has already been given elsewhere [2]. The present article [3] discusses the principle common to these systems, with particular emphasis on their new features in measurement and control technology. Attention will also be paid to the relationship between the frequency-analog technique and the familiar technique of frequency modulation, the effects of interference and the effects of changes in the waveform of the signals.

The signal parameters

In general the measured quantity x is of the analog type, which means that its possible values occupy a continuous band. In practice the conversion of x into a readable quantity z can never be made without some error. The result is that the value of z does not represent one x but rather a narrow band which in theory contains an infinite number of x-values. The total collection can therefore be considered to consist of a *finite* number M of these separately distinguishable bands of measured values. This means that at the receiving end

Dr.-Ing. D. Gossel is a Scientific Adviser with Philips Forschungslaboratorium Hamburg GmbH, Hamburg, West Germany.

we are only interested in M signal levels. The number of bits needed to indicate a measured value is given by:

$$N = \log_2 M. \quad (1)$$

If a noisy channel of bandwidth B is used for transmitting this quantity of information, we then have:

$$N = BT \log_2 (1 + P_s/P_n) = BTS. \quad (2)$$

Here T is the transmission time per quantity of information N, i.e. per measured value, while P_s is the average signal power and P_n the average noise power in the channel. The minimum time elapsing between the reception of two successive signals is therefore equal to T.

The quantity of information N can be represented in the form of a rectangular block whose sides have the lengths B, T and $S = \log_2 (1 + P_s/P_n)$ [4]. If a specified quantity of information is to be transmitted, only the volume of the block is then important. As long as equation (2) remains applicable, the sides of the block can be given arbitrary lengths.

An 'information block' of this kind is a useful device for choosing the various signal parameters. *Fig. 1a* shows the situation for complete amplitude-analog transmission. The entire quantity of information corresponding to an individual measured value is transmitted in this case in the form of *one* signal amplitude — and not in the form of a series of such amplitudes. The area of one of the sides of the block is then given by [5]:

$$B_1 T_1 = B_2 T_2 = \tfrac{1}{2}. \quad (3)$$

The restriction (3) means that the only remaining parameter is the number (M) of separately distinguishable bands; of course the possible signal-to-noise ratio in the transmission channel imposes a limitation on M.

Summarizing, in the case of amplitude-analog transmission it is only useful to improve the resolution (i.e. increase the number M) when determining the quantity x as long as the signal-to-noise ratio in the transmission channel is given a value in keeping with this increase. In practice a suitable transmission channel for values of M greater than between 100 and 1000 can be very expensive.

The situation becomes quite different if the transmission method permits a free choice of the product BT, for it is then possible to use an inexpensive transmission channel, with its given signal-to-noise ratio, and adapt the product BT in such a way as to give the information block the required volume (fig. 1*b*). The frequency or period can then be used as the signal parameter, and in this situation we thus have frequency-analog transmission. Frequency-analog signals, like digital signals, are not greatly affected by interference; in both cases this advantage arises because the magnitude of the signal amplitude has hardly any effect on the measured value. The information block must of course have the required volume, which means that the advantage referred to is obtained at the expense of either a larger bandwidth B or a longer measuring time T for a given measured value.

The transmission channels commonly used in process control often have a poor signal-to-noise ratio, but much of the available bandwidth is not used. If it were, a considerable improvement could be obtained, as we know from the theory of frequency modulation.

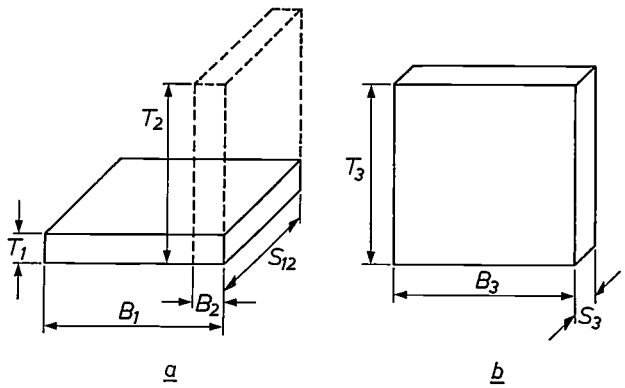


Fig. 1. Information blocks containing the same quantity of information. The three sides B, T and S, represent the three factors of the product of the right-hand side of equation (2). *a*) The situation in full amplitude-analog transmission ($B_1T_1 = B_2T_2 = \frac{1}{2}$). *b*) The situation in the case of transmission that is not amplitude-analog in nature, as in pulse-code modulation, for example. The product BT can be freely chosen.

[*] 'Unsere Forschung in Deutschland II', a joint publication of Philips Forschungslaboratorium Aachen and Philips Forschungslaboratorium Hamburg, published in the summer of 1973. The four articles referred to are:
G. Landvogt, Schwingende Blattfeder als frequenzanaloger Meßgrößenaufnehmer, pp. 218-221;
D. Meyer-Ebrecht, Converter für frequenzanaloge Meßsysteme, pp. 221-225 (see also Philips tech. Rev. **29**, 189-196, 1968);
M. Klinck, Ein digitaler Einzelregler für frequenzanaloge Meßwerte, pp. 225-227;
H. Kalis and J. Lemmrich, Frequenzanaloge Drehzahlregelungen, pp. 228-232 (see also Philips tech. Rev. **33**, 260-271, 1973).

[1] The term 'frequency-analog' was introduced as long ago as 1963 for certain types of automatic calculators; see H. Michaelis, Frequenzanalogierechner und digitaler Prozeßrechner, Regelungstechnik **11**, 114-118, 1963.

[2] D. Gossel, Meßsysteme und Regelungen mit Frequenzsignalen, Messen - Steuern - Regeln **14**, 22-28, 1971.

[3] A somewhat more detailed version of this article appeared in Elektrotechn. Z. A **93**, 577-581, 1972. The control engineering concepts and terms used in that article were in accordance with the German standard DIN 19226 (May 1968).

[4] K. Küpfmüller, Allgemeine Prinzipien der Nachrichtenübertragung, chapter 16, Kanalkapazität, in: H. Meinke and F. W. Gundlach, Taschenbuch der Hochfrequenztechnik, 3rd impression, Springer, Berlin 1968, p. 1420.

[5] See for example K. Küpfmüller, Einführung in die theoretische Elektrotechnik, 8th impression, Springer Berlin 1965, p. 381 or W. D. Hershberger, Principles of communication systems, Prentice-Hall, New York 1955, p. 46.

Frequency-analog signals

A good example of a frequency-analog signal is the output signal of a precision converter designed for use with strain-gauge measurements [6]. The frequency $\Omega/2\pi$ of the signal may vary from 200 to 1200 Hz; the smallest value corresponds to the smallest measured value x, and there is a very good linear relationship between the quantities Ω and x. To establish the link with the theory of frequency modulation we shall assume that the measured quantity x varies sinusoidally between two extreme values x_{min} and x_{max}, and that the angular frequency of this variation is ω_x (*fig. 2a*). The angular frequency Ω will then also vary sinusoidally about an average value Ω_0, which in our example is $2\pi \times 700$ radians per second (fig. 2*b*). The amplitude of the sinusoidal variation, here $2\pi \times 500$ rad/s, is denoted by $\Delta\Omega$, the frequency swing.

The instantaneous value of the angular frequency as a function of time is given by:

$$\Omega(t) = \Omega_0 + \Delta\Omega \sin \omega_x t = \Omega_0 (1 + \frac{\Delta\Omega}{\Omega_0} \sin \omega_x t), \quad (4)$$

where $\Delta\Omega/\Omega_0$ is the modulation factor. The phase $\Phi(t)$ is given by:

$$\Phi(t) = \int \Omega(t) \mathrm{d}t = \Omega_0 t - \frac{\Delta\Omega}{\omega_x} \cos \omega_x t + \text{constant}. \quad (5)$$

The ratio $\Delta\Omega/\omega_x$ is the modulation index η [7].

Table I lists some numerical values that are typical for the application of frequency-analog technique and of frequency modulation in radio transmissions. It can be seen that the modulation index — an important characteristic in the description of interference effects — is of the same order of magnitude in both cases. The modulation factor, on the other hand, is a thousand times smaller in frequency modulation. In applications of the new technique the unwanted effects of interference should be no greater than in frequency-modulated broadcast transmissions. Furthermore, the high modulation degree is an advantage in demodulation, i.e. converting the signal back into a readable signal of the amplitude-analog type.

The effects on FM signals of pulsed transients, noise and interfering signals which are themselves sinusoidal are known [7]. Provided the interference is not too great it appears in general that, compared with the situation for amplitude-analog type signals, the reduction in interference comes out at about the expected factor η. Of course, the exact value of the reduction factor will depend on the nature of the interfering effect. Introducing the parameters ϱ_A and ϱ_F, describing the relative interference level in the cases of amplitude-analog and frequency-analog signals, we can express

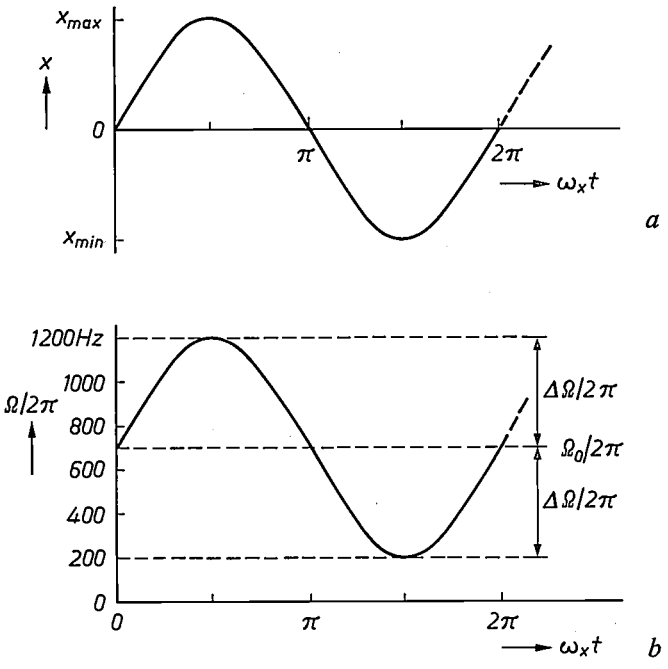


Fig. 2. Frequency-analog transmission of a sinusoidally varying quantity. *a*) The measured signal x as a function of time (in fact plotted against the number of radians); ω_x angular frequency of the measured quantity. *b*) The associated instantaneous value of the frequency $\Omega/2\pi$ of the transmission signal as a function of time. $\Delta\Omega/2\pi$ frequency swing. $\Omega_0/2\pi$ mean value.

Table I. Typical examples of the applications of frequency-analog technique and of FM modulation in radio transmissions. The symbols have the same significance as in fig. 2.

	Frequency-analog	Frequency modulation
$\omega_x/2\pi$	0.1 . . . 100 Hz	15 . . . 15 000 Hz
$\Delta\Omega/2\pi$	500 Hz	75 000 Hz
$\Delta\Omega/\omega_x = \eta$	5000 . . . 5	5000 . . . 5
$\Omega_0/2\pi$	700 Hz	100 MHz
$\Delta\Omega/\Omega_0$	7.15×10^{-1}	7.5×10^{-4}

this in the equation [8]:

$$\varrho_F = \varrho_A/\eta. \quad (6)$$

In the numerical examples (Table I) the reduction in interference achieved ranges from 1/5 to 1/5000.

So far we have assumed an ideal frequency demodulator that follows the instantaneous angular frequency $\Omega(t)$ continuously. A most important practical feature of frequency-analog signals is that they can readily be transformed into digital signals, provided a counter is used in the demodulation process. A known time base is then used to measure either the frequency or the period. The instantaneous angular frequency $\Omega(t)$ is not continuously followed for this purpose; instead, the zero-crossings of the frequency-analog signal are monitored. Nor is the read-out signal continuously available, but only at discrete instants in time, referred to as sampling times. The period of time between two successive sampling times is the 'sampling interval'. At

any given sampling time the read-out signal then represents a kind of average of all the frequencies $\Omega(t)$ that occurred during the preceding sampling interval. It amounts to saying that the transfer function contains an extra lowpass-filter [9] term

$$a_z = \frac{\sin \frac{1}{2}\omega_x T_0}{\frac{1}{2}\omega_x T_0}. \qquad (7)$$

It is then necessary to ensure that

$$T_0 < \frac{2\pi}{\omega_x} \quad \text{or} \quad \frac{\omega_x T_0}{\pi} < 2, \qquad (8)$$

since otherwise a_z could drop to zero. In accordance with a rule applicable to periodic sampling, it is necessary to satisfy the inequality

$$\frac{2\omega_x}{2\pi} < \frac{1}{T_0} \quad \text{or} \quad \frac{\omega_x T_0}{\pi} < 1. \qquad (9)$$

If (9) is satisfied, equation (8) is bound to be valid.

Processing as above with the numerical example $\omega_x/2\pi = 100$ Hz from Table I, and bearing in mind that T must not exceed a value such that, at a frequency equal to twice $\omega_x/2\pi$, a measurement is possible in every half-cycle [10], or $T_0 \leqslant 1/2\times 200$, we then find:

$$\frac{\omega_x T_0}{\pi} = \frac{2\pi\times 100}{\pi}\,\frac{1}{2\times 200} = \tfrac{1}{2},$$

so that (9) is in fact satisfied.

The limitation to sampling at the zero crossings results in a greater reduction in the interference level than follows from equation (6). This is mainly due to the less frequent occurrence of interfering effects and not so much to a reduction in their magnitude. At the zero crossings the only interference that produces any effect is that which causes one zero crossing too many or which, being close to the crossing, causes a shift in the time of occurrence of the zero crossing. In the case of amplitude-analog signals, and also with *continously* followed frequency-analog signals, *every* interfering signal will in fact affect the output signal.

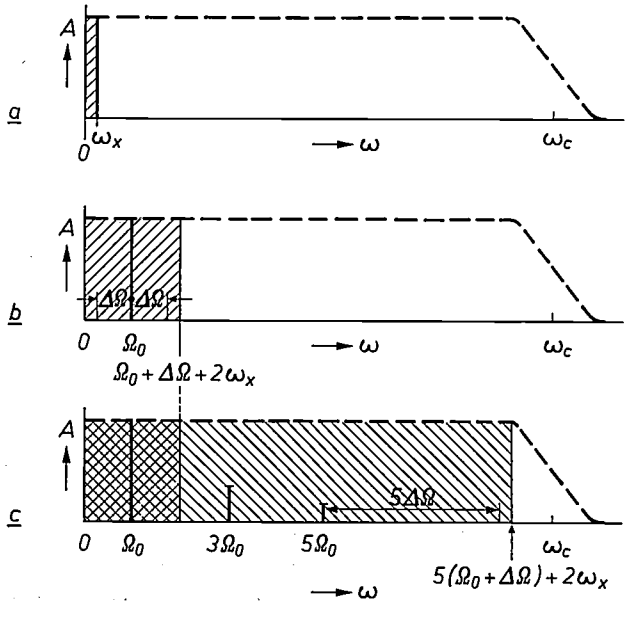


Fig. 3. The bandwidth required: *a*) for an amplitude-analog signal; *b*) for a sinusoidal frequency-analog signal; *c*) for a frequency-analog signal consisting of rectangular pulses. A amplitude. ω angular frequency. For ω_x, $\Delta\Omega$ and Ω_0, see fig. 2.

It is not so simple to make general statements about the extra suppression of interference resulting from the limitation of sampling to the zero crossings. Many different factors are involved, such as the waveform of the signal, the ratio of the instantaneous values of signal and interference, and also the ratio of the duration τ_n of an interfering pulse and the period τ_s of the useful signal. If the useful signal consists of rectangular pulses and if there are no extremely short interfering pulses, the effect on the output signals, at least on their frequency, will be reduced by a factor of

$$\frac{\tau_n}{\frac{1}{2}\tau_s} = \frac{2\Omega}{\omega_c}. \qquad (10)$$

Here ω_c is equal to 2π times the upper frequency limit of the transmission channel. In our example, where $\Omega/2\pi = 1200$ Hz and $\omega_c/2\pi = 1.2$ MHz, the interference is thus further reduced by a factor of 500.

Many frequency-analog converters do indeed provide signals consisting of rectangular pulses instead of sinusoidal signals, which means that appreciably more bandwidth is needed. *Fig. 3a* shows the bandwidth required for an amplitude-analog signal; figs. 3*b* and *c* illustrate the cases of a frequency-analog signal, one sinusoidal (*b*) and the other rectangular (*c*). In the latter case the signal in the transmission channel can be approximated by:

$$y_t = \frac{4}{\pi} H_0 (\sin \Phi + \tfrac{1}{3}\sin 3\Phi + \ldots + \frac{1}{k}\sin k\Phi), \qquad (11)$$

where k is a positive odd integer and H_0 is the amplitude of the rectangular pulses; $\Phi(t) = \Omega_0 t - \eta \cos \omega_x t$ (see eq. 5).

In the case where $\Omega_0/2\pi = 700$ Hz, $\Delta\Omega/2\pi = 500$ Hz and $\omega_x/2\pi = 100$ Hz we therefore require a bandwidth of 1400 Hz for a sinusoidal signal ($k = 1$), and a bandwidth of 11 000 Hz for a signal consisting of rectangular pulses, if the highest harmonic contained in the

[6] See the article by D. Meyer-Ebrecht [*].
[7] K. Küpfmüller, Die Systemtheorie der elektrischen Nachrichtenübertragung, 3rd impression, Hirzel, Stuttgart 1968.
[8] The relative interference level introduced here is simply the reciprocal signal-to-noise ratio; a derivation of (6) will be found in the article of note [3].
[9] G. Landvogt and D. Meyer-Ebrecht, Frequency analogy, a powerful tool for process instrumentation, Proc. 5th IFAC Congress, Paris 1972.
[10] See the article of note [3].

signal is the ninth ($k = 9$). The information is thus transmitted by the rectangular-pulse signal in the same way as it would be by the sum of k sinusoidal FM signals. The kth harmonic of the signal has an amplitude equal to $1/k$ times that of the fundamental, whereas the modulation index is k times greater.

Closer analysis of the situation in which the interfering pulses are very much shorter in duration than the rectangular pulses has revealed that the relative level of interference is about ten per cent higher than in the case of a sinusoidal FM signal [11]. This result depends on the number of harmonics in the rectangular-pulse signal, provided the signal is not too small. The larger bandwidth needed for rectangular pulses does *not* reduce the level of interference. It therefore seems better to convert the rectangular-pulse signal into a sinusoidal signal, if the transmission channel has a large bandwidth. The sinusoidal signal should then have a large frequency swing or modulation index to correspond to the bandwidth, since the reduction of interference increases in proportion to the modulation index η. This implies that the reduction is approximately proportional to the bandwidth.

If the rectangular-pulse signal does not pass through a filter before transmission, the frequencies it contains will be limited by the transmission channel itself. In such a case the sidebands belonging to the higher harmonics of the carrier will not be symmetrically transmitted. The transmission will consequently be subjected to an additional error. An investigation of the behaviour of the transmission in the time domain has shown that this error also depends on the filter characteristics of the transmission channel [9].

Characteristics in measurement and control technology

In measurement techniques instruments that operate with frequency-analog signals have the following good features.

— There is no restriction on the choice of the signal amplitude — which does *not* depend on the measured quantity. The amplitude and the transmission channel used can be optimally matched to one another. If, for example, an amplitude of 10 volts is chosen, the signal level will be far higher than most of the interference encountered.

— In any transmission channel the measurement time can easily be adapted to the level of interference present, so that the measuring accuracy required can always be achieved. If, for example, the zero crossings cannot be determined accurately enough because of strong interference, all that is necessary is to take the measured value from a large number of periods instead of just one.

— The transmission channels required work with a.c. currents. Existing carrier-wave equipment or ordinary telephone lines can be used. Just as with the telephone, the same line serves for the transfer of energy and information at the same time.

— Amplifier drift, or thermal and galvanic effects at contact points and in switches have no adverse effect.

— Unlike the conventional amplitude-analog-to-digital conversion, frequency-analog-to-digital is very simple. It is done with counter circuits, and with simple means it is nevertheless possible to achieve very high accuracy with this technique. It is also easy to connect to existing digital equipment for data processing.

— The conversion into an amplitude-analog signal amounts to the production of a d.c. signal, which is produced as an average of pulses covering the same area; their magnitude is proportional to the frequency of the transmission signal. The equipment is thus compatible with conventional recording instruments such as pen recorders.

— Continuous counting of the periods — by digital means — also enables the integral of a measured quantity to be determined. This determination is made with absolute accuracy; there is no systematic deviation and no drift. In this way, a quantity can be determined from a flow rate, a phase angle from a speed of revolution, or a path-length from a velocity.

— It is easy to determine at any instant whether an installation working with frequency-analog signals is ready to be put into operation. The operator only needs headphones, provided at least there are signals available in the audible frequency range. The pitch of the audible tone can also be an indication of the value of the measured quantity at that instant.

In frequency-analog control systems the measured values are available in the form of frequencies, but the desired value of the controlled condition and possible auxiliary quantities are available in digital form [11].

A complete controller in such a system, or individual elements of it, can readily be made in integrated-circuit form. Controllers of this type can be an attractive proposition in the following circumstances:

— When extreme accuracy is necessary in the manufacture of a product (e.g. for controlling mix ratios in the chemical industry).

— When calibration is required.

— When the differences between almost identical quantities have to be determined (e.g. in the case of a mass or energy balance).

— When movements in machines have to be synchronized (e.g. speed ratios in presses, paper machines, machines in the textile industry).

— When the control system is to be incorporated in a digital signal-processing installation.

A comparison of our new, individual controllers with an automatic calculator that controls a large number of processes in time-division multiplexing (known as Direct Digital Control or DDC) shows that our devices have the following advantages:

— They can work economically even when the number of control loops involved is small.

— They do not require special sampling units or buffer stores.

— They monitor processes continuously.

— The control loops are completely isolated from each other.

— The installations can easily be extended.

On the other hand these individual controllers can if required, easily be put under central automatic supervisory control, which can adjust the controls if changes occur in parameters that define their control ranges. If the automatic system should fail, the installation as a whole nevertheless goes on working, since everything in the control loops continues to operate on the basis of the last available values.

A slight disadvantage of the new technique is that accurate converters are not yet available for all quantities that one would like to measure. In addition the investment in equipment is still relatively expensive, especially for applications in which there are no particular requirements for high accuracy and freedom from interference. Nowadays, however, with low-cost integrated circuits, this particular point is of diminishing significance. It is safe to say that in the near future the amplitude-analog technique will continue to exist side by side with the frequency-analog technique, each with their own areas of application.

Summary. Transducers based on frequency-analog parameters (frequency or period) of an a.c. signal, instead of amplitude-analog parameters such as currents, voltages, resistances or air pressures, are very much less sensitive to interference. This is a particular advantage in digital data processing. The article discusses the characteristics of the frequency-analog measuring technique in the general context of measurement and control. Emphasis is placed on the connection between frequency modulation and frequency-analog techniques, on the behaviour of frequency-analog systems in strong interference, and on the effect of changes in the waveform of the periodic signals.

[11] See for example the articles by H. Kalis and J. Lemmrich and the one by M. Klinck [*].

Recent scientific publications

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

Philips Research Laboratories, Eindhoven, 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.

S. Arnemann & M. Tasto: Generating halftone pictures on graphic computer terminals using run length coding.
Computer Graph. Image Proc. **2**, 1-11, 1973 (No. 1). *H*

P. Blood & J. W. Orton: The electrical properties of *n*-type epitaxial InP in the temperature range 5 K to 700 K.
J. Physics C **7**, 893-904, 1974 (No. 5). *M*

J. Brokken-Zijp & H. v.d. Bogaert: Influence of solvent and temperature on decomposition and isomerisation of Z-alkylaryldiazo sulphides.
Tetrahedron **29**, 4169-4174, 1973 (No. 24). *E*

G. Brouwer, Th. P. M. Meeuwsen & A. W. Witmer: A rapid method for obtaining optimum focusing in spark-source mass spectrometers with photographic plates.
Int. J. Mass Spectrom. Ion Phys. **12**, 397-402, 1973 (No. 5). *E*

H. G. Bruijuing, W. J. Kleuters & P. J. Poolman: Ignition and electronic injection control for the future.
Proc. Conf. on Passenger car engines, London 1973 (Instn. Mech. Engrs. Conf. Publ. 19), pp. 126-132. *E*

P. C. M. N. Bruijs (Philips Lighting Division, Eindhoven): The determination of rare earths in yttrium oxide by means of a preconcentration technique and Ge(Li) gamma-spectrometry.
J. radioanal. Chem. **16**, 115-122, 1973 (No. 1).

T. A. C. M. Claasen, W. F. G. Mecklenbräuker & J. B. H. Peek: Second-order digital filter with only one magnitude-truncation quantiser and having practically no limit cycles.
Electronics Letters **9**, 531-532, 1973 (No. 22). *E*

C. D. Corbey: The characteristics of Impatt diodes in relation to wideband varactor tuned oscillators.
Acta Electronica **17**, 187-192, 1974 (No. 2). *M*

J. Daniels: Calculation of the dielectric constants of ferroelectric ceramics.
Phys. Stat. sol. (a) **21**, 497-505, 1974 (No. 2). *A*

P. Delsarte: Four fundamental parameters of a code and their combinatorial significance.
Information and Control **23**, 407-438, 1973 (No. 5). *B*

A. M. van Diepen & R. P. van Stapele: Ordered local distortions in cubic $FeCr_2S_4$.
Solid State Comm. **13**, 1651-1653, 1973 (No. 10). *E*

J. Donjon, F. Dumont, M. Grenot, J.-P. Hazan, G. Marie & J. Pergrale: A Pockels-effect light valve: Phototitus. Applications to optical image processing.
IEEE Trans. **ED-20**, 1037-1042, 1973 (No. 11). *L*

J. W. F. Dorleijn & W. F. Druyvesteyn: Stability of bubbles subjected to an in-plane field.
IEEE Trans. **MAG-9**, 521-522, 1973 (No. 3). *E*

G. Eschard, J. Graf & R. Polaert: Les intensificateurs d'images à microcanaux pour la détection à faible niveau lumineux.
Onde électr. **53**, 255-260, 1973 (No. 7). *L*

G. Eschard, A. Pélissier, J. Paulin, J. F. Bonnal & H. Bernardet (SODERN): The engineering model of a cesium contact ion thrustor: design and development of the thrustor module.
Proc. Conf. on Electric propulsion of space vehicles, Abingdon 1973, pp. 56-60. *L*

A. Farrayre & B. Kramer: Réalisation et caractérisation de diodes à avalanche en GaAs fiables et reproductibles.
Acta Electronica **17**, 99-113, 1974 (No. 2). *L*

A. Farrayre & A. Mircea: Distribution latérale du courant et de la température dans les dispositifs hyperfréquences.
Acta Electronica **17**, 115-125, 1974 (No. 2). *L*

R. C. French & R. F. Mitchell: Class of nonorthogonal transformations for signal processing.
Electronics Letters **10**, 78-79, 1974 (No. 6). *M*

A. A. van der Giessen: A magnetic recording tape based on iron particles.
IEEE Trans. **MAG-9**, 191-194, 1973 (No. 3). *E*

J. J. Goedbloed & M. T. Vlaardingerbroek: Noise in Impatt-diode oscillators.
Acta Electronica **17**, 151-163, 1974 (No. 2). *E*

J. M. Goethals: Le contrôle des erreurs dans les transmissions digitales.
Automatisme **19**, 17-24, 1974 (No. 1). *B*

J. C. M. Henning & J. H. den Boef: Strain modulated electron spin resonance (SMESR) of Cr^{3+} in MgO: evidence for induced off-diagonal elements in the *g*-tensor.
Physics Letters **46A**, 183-184, 1973 (No. 3). *E*

S. van Houten: Display devices.
Solid State Devices 1973, Proc. 3rd Eur. Conf., Munich, pp. 131-157; 1974. *E*

A. P. Hulst: On a family of high-power transducers.
Ultrasonics International 1973, Proc. Conf. London, pp. 285-294. *E*

W. H. de Jeu: First- and second-order nematic-smectic *A* phase transitions in the series of di-*n*-alkyl azoxybenzenes.
Solid State Comm. **13**, 1521-1523, 1973 (No. 9). *E*

F. A. de Jonge, J. A. L. Potgiesser, D. L. A. Tjaden & U. Enz: Recording with magnetic bubbles.
IEEE Trans. **MAG-9**, 179-182, 1973 (No. 3). *E*

E. Krätzig: Ultrasonic investigation of the superconducting surface sheath.
Phys. Stat. sol. (b) **60**, K 79-81, 1973 (No. 2). *H*

J.-P. Krumme, W. Tolksdorf, H. Dimigen & H. Hieber: Formation of arbitrary compensation temperature profiles in epitaxial ferrimagnetic garnet films.
Phys. Stat. sol. (a) **20**, 725-729, 1973 (No. 2). *H*

T. P. R. Linnecar: Maximum tolerable local oscillator noise in a 12-GHz f.m. satellite television receiver.
Radio & Electronic Engr. **44**, 77-84, 1974 (No. 2). *M*

R. Metselaar & M. A. H. Huyberts: The stoichiometry and defect structure of yttrium iron garnet and the nature of the centres active in the photomagnetic effect.
J. Phys. Chem. Solids **34**, 2257-2263, 1973 (No. 12). *E*

A. R. Miedema & M. H. van Maaren: Superconductivity in ternary alloys of transition metals.
Physica **69**, 308-316, 1973 (No. 1). *E*

A. Mircea & R. Perichon (Université de Lille I): Origines et mécanismes du bruit de fond dans les diodes à avalanche et à temps de transit.
Acta Electronica **17**, 165-170, 1974 (No. 2). *L*

T. G. J. van Oirschot & W. Nijman: Improved boat for multiple-bin liquid phase epitaxy.
J. Crystal Growth **20**, 301-305, 1973 (No. 4). *E*

J. A. Pals & W. J. J. A. van Heck: Peaked structure in field-effect mobility of silicon MOS transistors at very low temperatures.
Appl. Phys. Letters **23**, 550-552, 1973 (No. 10). *E*

K. Pape & H. Hieber: The electron microscopic investigation of sputtered and electrolytically deposited thin magnetic layers.
Pract. Metallogr. **11**, 10-18, 1974 (No. 1). *(Also in German.)* *H*

A. Pirotte: Automatic theorem proving based on resolution.
Ann. Rev. aut. Progr. **7**, 201-266, 1973 (No. 4). *B*

P.-A. Rolland, E. Constant, A. Derycke (all with Université de Lille I) **& J. Michel:** Multiplication de fréquence par diode à avalanche en ondes millimétriques.
Acta Electronica **17**, 213-228, 1974 (No. 2). *L*

A. Semichon: Mise en œuvre des diodes à avalanche pour hyperfréquences.
Acta Electronica **17**, 171-180, 1974 (No. 2). *L*

G. Simon (Technische Universität Braunschweig) **& G. R. Zeller:** Coulomb energy of cubic lattices.
J. Phys. Chem. Solids **35**, 187-194, 1974 (No. 2). *A*

B. M. Singer & J. Kostelec: Theory, design, and performance of low-blooming silicon diode array imaging targets.
IEEE Trans. **ED-21**, 84-89, 1974 (No. 1). *N*

A. M. J. Spruijt: The twist-disclination line in planar oriented samples of liquid crystals.
Solid State Comm. **13**, 1919-1922, 1973 (No. 11). *E*

F. L. H. M. Stumpers: The 1973 CISPR Plenary Assembly at Monmouth College, New Jersey.
IEEE Trans. **EMC-15**, 197-199, 1973 (No. 4). *E*

E. J. Tercic: Superposition measurements with a flux-sensitive head in digital magnetic recording.
IEEE Trans. **MAG-9**, 335-338, 1973 (No. 3). *E*

D. L. A. Tjaden: Some notes on 'superposition' in digital magnetic recording.
IEEE Trans. **MAG-9**, 331-335, 1973 (No. 3). *E*

H. Tjassens: Circuit analysis of a stable and low noise Impatt-diode oscillator for X-band.
Acta Electronica **17**, 181-185, 1974 (No. 2). *E*

M. L. Verheijke & J. C. Verplanke: Overall instrumental thermal neutron activation analysis of high-purity materials.
J. radioanal. Chem. **15**, 509-515, 1973 (No. 2). *E*

J. C. Verplanke & P. N. Kuin: Tracer study of a chemical separation procedure for the determination of impurities in selenium by means of thermal neutron activation analysis.
J. radioanal. Chem. **16**, 57-66, 1973 (No. 1). *E*

S. Wittekoek & T. J. A. Popma: Magneto-optic Kerr rotation of bismuth-substituted iron garnets in the 2-5.2-eV spectral range.
J. appl. Phys. **44**, 5560-5566, 1973 (No. 12). *E*

Contents of Philips Telecommunication Review **32**, No. 3, 1974:

R. H. Baylis & S. Brcic: PHL 7404 ILS electronic modulator (pp. 93-104).
F. L. van den Berg & J. Mulder: The Philips VOR beacon type RN 100 (pp. 105-116).
E. C. Priebee: SARP air traffic control system (pp. 117-127).
A. H. Brands: The random access bright display (pp. 128-139).
E. C. Priebee: Radar systems for air traffic control (pp. 140-154).
S. J. Robinson: MADGE, a portable aircraft landing aid (pp. 155-167).
C. J. Aangeenbrug & P. Bikker: VHF AM transmitting equipment for ground-to-air communication (pp. 168-178).
M. V. Callendar & W. P. Graville: Automatic VHF DF equipment type CE 254 (pp. 180-183).
J. P. Landrot: FM/CW radio altimeters (pp. 184-191).
V. J. Cox & P. G. Dowty: Airborne tactical radar ARI 5955 and Transponder ARI 5954 (pp. 192-200).
P. L. Stride & V. J. Cox: Airborne weather radar (pp. 201-211).

Contents of Valvo Berichte **18**, No. 1/2, 1974 *(Sonderheft zum 50jährigen Firmenjubiläum)*:

R. Suhrmann: Neue Schaltungssysteme für Farbfernsehgeräte (pp. 5-12).
Ein PAL-Decoder mit drei integrierten Schaltungen (pp. 13-58):
E. Pech: Eine integrierte Synchrondemodulatorkombination für Farbfernsehempfänger (pp. 15-28),
K.-H. Mathies: Weiterentwicklung einer integrierten Farbartkombination für Farbfernsehempfänger (pp. 29-46),
K. Juhnke: Weiterentwicklung einer integrierten Leuchtdichtekombination für Farbfernsehempfänger (pp. 47-58).
H. H. Feindt: Die Entwicklung der integrierten Horizontalkombination V 480 (pp. 59-76).
L. Grimm, K.-J. Hilke, E. Scharrer & U. Schlenker: Physikalisch-chemische und chemische Vorgänge bei der Herstellung von Farbbildröhren, 1. Teil (pp. 77-93).
K. Böke & W. Reichardt: Elektronenoptik in Fernsehbildröhren (pp. 95-100).
W. Aschermann: Entwicklungstendenzen bei Rundfunk- und Phonogeräten (pp. 101-107).
B. P. Bahnsen: Integrierte Schaltungen im Rundfunk-Empfangsteil (pp. 109-120).
E. A. Kilian: Ein neues Konzept für den Wiedergabeteil von Rundfunk- und Phonogeräten (pp. 121-127).
U. Schillhof: Bedienungserleichterung für Rundfunkgeräte unter besonderer Beachtung der Empfängerabstimmung (pp. 129-142).
R. Ranfft: Schalt-Netzteile mit schnellen Transistoren hoher Sperrspannung (pp. 143-154).
M. Herrmann: Vakuumlose elektronische Bildaufnahmeeinheit mit 32×32 Bildelementen (pp. 155-160).
W. Schmidt: Zur Systematik und Anwendung von Leistungs-Mikrowellenröhren in der heutigen Nachrichten- und Funkortungstechnik (pp. 161-167).
W. Schmidt: VALVO-Leistungsklystrons (pp. 169-179).
E. Ginsberg: Typenentwicklung monolithisch integrierter Schaltungen (pp. 181-188).
E. Bauböck & G. Renelt: Rechnerunterstützter Entwurf von Layouts integrierter Schaltungen (pp. 189-202).
H. Dammann: Eine mit kohärentem Licht arbeitende Filtermethode für Masken mit periodischer Struktur (pp. 203-214).
O. Jakits: Integrierte Injektionslogik, ein neuartiges Prinzip für Digitalschaltungen (pp. 215-226).
D. Eckstein: Ionenimplantation als Dotierverfahren in der Halbleitertechnologie — Eine Übersicht (pp. 227-234).
E. Uden: Dual-in-line-Gehäuse mit 16 Anschlüssen für integrierte Halbleiterschaltungen verschiedener Verlustleistungen (pp. 235-242).
M. Lemke & W. Schilz: Breitband-Richtungsleitungen bis 18 GHz (pp. 243-250).
W. Laurich & H. Wieters: Eigenschaften von gedruckten Kondensatoren und Spulen in Dickschichttechnik (pp. 251-266).
F. Rott: Biegeelemente aus dem piezoelektrischen Werkstoff PXE (pp. 267-276).
L. Eiermann: Ferroxcube 3 C 8, ein Transformatorkernmaterial mit vielseitigen Anwendungsmöglichkeiten (pp. 277-287).

PHILIPS TECHNICAL REVIEW

VOLUME 34, 1974, No. 11/12

Inorganic chemical analysis

The special issue completing this volume of Philips Technical Review is given over entirely to chemical analysis, as practised today. To keep the size of the issue within reasonable bounds we have not included any articles on organic analysis. However, one 'border-line case' has been included: the study of surfaces and surface layers by bombardment with photons, electrons or ions, sometimes in the form of a thin beam.

The most significant development in inorganic chemical analysis during the last 30 years has been — and still is — the success of methods based on physical phenomena. Nevertheless, it would be wrong to assume that these new methods have overtaken the classical methods of chemical analysis. In accuracy, for example, some of the classical methods are still unsurpassed. Because of the large number and great variety of the methods now available, the correct approach to an analysis is nowadays often a complicated affair, in which many factors — time required, cost, accuracy, etc. — all have to be taken into account.

The more widespread use of instruments and the growth of the market have highlighted the contribution from a group who previously played only a minor part in analytical chemistry — those who develop and manufacture the instruments. Although Philips occupy a leading position here, this work will only be mentioned incidentally in this issue.

The introductory article gives an account of the general pattern of the work at the Philips research laboratories. The other articles in this issue include two that to some extent are of a survey nature — on atomic spectrography and surface studies — and also a number of shorter contributions mainly concerned with recent improvements in instrumental methods of analysis.

Philips tech. Rev. 34, 298-304, 1974, No. 11/12

Inorganic chemical analysis

W. F. Knippenberg

Chemical analysis today

Although chemical analysis is a branch of applied science that might not be strong in imaginative appeal, it is of considerable importance to humanity in many ways. It has become an indispensable aid in many areas of scientific and technological research, in medicine, in the control of foodstuffs and in environmental control. Analytical chemistry serves both for identifying and characterizing substances and for studying or following a wide variety of processes that involve the conversion of materials. These substances and processes may be either of natural origin or man-made. Analytical chemistry includes both the performance of analyses and the investigation of methods. Since it is an applied science, its development is determined largely by changes in the problems that analytical chemists are called upon to solve and by advances made in the basic sciences and in technology.

The object of an *inorganic* chemical analysis in its most general form is to establish the chemical composition of an inorganic substance; in other words, to determine the elements contained in a given inorganic substance (qualitative analysis) and the amounts in which they are present (quantitative analysis). It may also be desirable to know the manner in which the elements are combined in groups, in molecules or in separate phases, and if so, how these phases are distributed and whether their composition is homogeneous. In practice, however, an analysis is usually carried out to provide answers to a limited number of specific questions. The ability to answer these questions depends to a great extent on whether material is available in relatively large quantities (macroanalysis, 10-100 mg), small quantities (microanalysis, 1-10 μg; ultramicroanalysis, ng-μg) and whether destructive analysis of the sample is permissible.

Examples of situations in which a limited number of questions have to be answered are to be found in materials research, environmental control and process monitoring and control.

In investigations of technical materials, substances whose composition is broadly known often have to be analysed to provide an accurate quantitative determination of the major constituents (more than 100 mg/g), or a somewhat less accurate determination of the minor constituents (10-100 mg/g), admixtures (10 mg/g - 100 μg/g), trace elements (1-100 μg/g) and ultratraces (ng/g - μg/g).

In environmental control the object is to determine trace elements and ultratraces in air, water and soil.

In process monitoring it is necessary to follow the variations in the concentration of a particular substance or, for example where a tracer is used, to determine one element accurately in a series of samples selected in time or location.

Questions of molecular composition arise, for example, in the analysis of organometallic compounds. The methods of *organic* chemical analysis, which has its own characteristic problems to solve, are to some extent applied in the analysis of such compounds. Whereas inorganic chemical analysis is concerned with the separation and identification of *elements*, organic chemical analysis deals with the separation and identification of *molecules*. Because of isomerism, a determination of elements is not sufficient for the purposes of organic chemistry, and the determination of molecular structure is essential here. Organic chemical analysis has developed its own characteristic methods of doing this, which include chromatography, mass spectrometry, molecular spectroscopy and nuclear magnetic resonance spectrometry.

Problems resembling those encountered in organic chemical analysis also feature strongly in the techniques for identifying molecules in inorganic gases and liquids. Comparable problems also arise in the analysis of solids when the chemist wishes to determine not only the nature of the elements but also the valency or the charge of the elements, or the 'functional groups' in which certain elements occur. Since the procedure of making a substance accessible to analysis, for example by solution, can affect the valency of an element and consequently the composition of an atomic group or radical, a combined chemical and physical approach has often been indispensable for these determinations. To characterize atomic groups in solids the inorganic analytical chemist often has to resort to X-ray structural analysis, and for the determination of charge he must turn to some form of nuclear or electron spectrometry.

Dr W. F. Knippenberg is with Philips Research Laboratories, Eindhoven.

After a mechanical or chemical phase separation it is sometimes useful to carry out an X-ray structural analysis to check the nature and 'purity' of the solid phases being analysed, or to reach a conclusion about polymorphism. A homogeneity check of this type is important, because if more than one phase is present the results of the analysis could lead to a wrong interpretation, for example of the stoichiometric composition. If this has to be determined with an accuracy better than that of the phase determination by X-ray diffraction (1 % absolute) it is necessary to adopt other methods of verifying the homogeneity, such as the electron microprobe, or perhaps to carry out a microanalysis instead of a macroanalysis.

Investigation of analytical methods

Investigations in inorganic chemical analysis, can broadly be divided into three areas, corresponding to three groups of analysts. In the first place there is the large group of analysts who are primary engaged in carrying out analyses by conventional methods. They make no direct contribution to the development of new methods or of equipment, but they do help to improve the standard methods available by continually checking their results against those found by other methods, making it possible, for instance, to track down sources of systematic errors.

The second group consists of investigators working exclusively on the improvement of methods or looking for entirely new ones. Their investigations are usually prompted by analytical problems for which no entirely satisfactory solution can be found with existing methods. We shall return to this in more detail presently.

The third group consists of chemists mainly engaged in the design and refinement of analytical equipment.

The main trends in current analytical developments are towards greater dynamic range (the mass or concentration interval in which methods can be used), improved accuracy, shorter analysis times and increased automation. Since there are limits to the extent in which any given method can be improved in all these ways, the search for methods based on previously untried principles is very important.

The two main trends in the development of equipment are in opposite directions: one towards small instruments capable of providing a reliable answer to one particular question, and the other towards the 'complete analytical machine'.

After the analytical chemist has been asked to analyse a particular sample in a certain respect, he usually has to pretreat the sample before performing the analysis. This preparatory treatment comprises operations that should provide an analysis sample that can optimally provide the information that the client requires and will also meet the requirements imposed by the selected method of analysis. Research in this field is under way, and includes studies on methods for controlled grinding, dissolution or dispersion of solids, for phase separation (extraction and concentration), etc. With the continuous refinement and automation of the methods of analysis this kind of study is gradually assuming a greater importance.

The bewildering variety of analytical methods

The modern analytical chemist has an impressive array of methods available to him. Some idea of their variety can be obtained from *Table I*, which summarizes the methods available at Philips Research Laboratories for carrying out analyses as a service within the company. Attempts have been made in the past to classify the methods used in chemical analysis by distinguishing, for example, between instrumental and classical chemical methods, physical and chemical methods, methods of separating signals and methods of separating components, etc. In the last twenty or thirty years, however, instruments, whether simple or complex, have invaded analytical chemistry on such a scale that classifications of this type do not make much sense, since a clear dividing line can no longer be drawn between the different areas of analysis.

One distinction still worth making is based on the difference in approach to the analytical problem: whether the analysis consists of a series of single-element determinations or of a simultaneous multielement determination. Single-element analysis follows on from the classical chemical tradition, aiming at the selective isolation of a single element by the separation of components. Single-element determination may be done either instrumentally or noninstrumentally, and the method used may be either physical or chemical. A simultaneous multielement analysis, on the other hand, is always performed with instruments and physical methods. The instrument usually has to be calibrated by means of synthetic reference samples or with the aid of reference samples obtained by accurate analysis of existing samples with a single-element method.

The important features of the various methods for the determination of a particular element, such as sensitivity, the mass and concentration detection limit, reproducibility and accuracy, vary unsystematically. As Table I shows, however, the methods can to some extent be classified in terms of area of application since the values of the quantities mentioned above usually vary within definite limits. In practice, the choice of a method for a given analysis may be dictated by interelement effects, sensitivity to interference, cost, the time required for an analysis and other incidental factors. To obtain all the information desired it may

Table I. Methods of survey analysis and of single-element and multielement determinations as used in Philips Research Laboratories.

Method	Sample (in descending order: form, normal quantity and minimum quantity)	Application (number of elements between brackets)	Concentration or mass detection limit for majority of elements	Reproducibility (standard deviation)	Accuracy (possible deviation from real value)
Gravimetry	solutions 10^{-1}-10^{-2} % 100 ml 5 ml	single-element determination (30) of major and minor constituents		10^{-3}-10^{-1} %	10^{-2}-1 %
Volumetry	solutions 10^{-1}-10^{-2} % 100 ml 5 ml	single-element determination (60) of major and minor constituents		10^{-2}-1 %	0.5-1 %
Coulometry	solutions 10^{-1}-10^{-4} % 10 ml 0.1 ml	single-element determination (10) of major and minor constituents	0.1 μg	10^{-2}-1 %	10^{-1}-1 %
Polarography	solutions 10^{-4}-10^{-8} % 10 ml 1 ml	single-element determination (20) in composite material with 3-5 elements; trace elements, ultratraces	ng/g-μg/g	10 %	25 %
Spectrophotometry	solutions 10^{-1}-10^{-4} % 10 ml 1 ml	single-element determination (40) in composite material; of admixtures to trace elements	1 μg/g	0.5 %	1 %
Atomic spectroscopy (without chemical separation): Flame emission Flame absorption Flame fluorescence Furnace absorption	solutions 10^{-3}-10^{-6} % 5 ml 1 ml (furnace 1 μl)	single-element determination (70) in composite material with 15-20 elements; major constituents to ultratrace	0.1 μg/g 10^{-12}-10^{-11} g	1-10 %	factor of 2 (with chemical separation 10 %)
Emission spectroscopy using d.c. arc	solid 1-10 mg 0.5 mg	survey analysis for 69 elements; major constituents to trace elements	1-100 μg/g	10 %	factor of 3 with reference samples: 10 %
Emission spectroscopy with r.f. argon plasma	solutions 1-10^{-6} % 5 ml 100 μl	multielement determination (69) in composite material with 5-10 elements; major constituents to trace elements	0.01-1 μg/g 10^{-13}-10^{-11} g	2 %	with reference samples: 5 %
X-ray fluorescence spectroscopy	solid or solution layer: thickness about 1 mm, cross-section 30 mm	non-destructive multielement determination in composite material with 10-20 elements; major constituents to trace elements	1-10 μg/g	0.1-1 %	with reference samples: 0.5 %
Spark-source mass spectroscopy	solid 10-50 mg 1 μg (on substrate)	survey analysis for 85 elements; trace elements, ultratraces	10-100 ng/g	10 %	factor of 3 with reference samples: 20 %
Neutron activation analysis (gamma spectrometry)	solid or solution 1 g 1 mg	survey analysis for 60 elements; trace elements, ultratraces	ng/g-pg/g	10 %	20-50 %
Solid or solution		single-element determination (special conditions)		0.1-1 %	with reference samples: 0.1-1 %
Electron microprobe (EMA)	solid thin film: thickness 1-5 μm, cross-section 1-3 μm	survey analysis for 80 elements; major constituents to admixtures	10 mg/g-100 μg/g	2 %	5-10 %

Method	Sample (in descending order: form, normal quantity and minimum quantity)	Application (number of elements between brackets)	Concentration for mass detection limit for majority of elements	Reproducibility (standard deviation)	Accuracy (possible deviation from real value)
Auger electron spectroscopy (AES)	surface layer cross-section 10 μm-10 mm thickness 2-10 atomic layers, and profile analysis by sputtering - μm	survey analysis, except for H and He; major constituents to admixtures	10^{-3} monolayer	2%	factor of 2 with reference samples: 10%
Ion scattering spectrometry (ISS)	surface layer cross-section 50 μm-10 mm thickness 1 monolayer, and profile analysis by sputtering - μm	survey analysis; major constituents, minor constituents, trace elements	10^{-5}-10^{-3} monolayer	1%	factor of 2 with reference samples: 10%
Ion-induced photo emission (IIPE)	surface layer cross-section 1-10 mm thickness 1 monolayer, and profile analysis by sputtering - μm	survey analysis for 60 elements; major constituents, minor constituents, trace elements	10^{-6}-10^{-2} monolayer	2%	factor of 2 with reference samples: 10%
Secondary-ion mass spectrometry (SIMS)	surface layer cross-section 1 μm-3 mm thickness 1 monolayer, and profile analysis by sputtering - μm	survey analysis (all elements from thickness of 40 nm (400 Å), one element from thickness 0.3 nm (3 Å); major constituents, minor constituents, trace elements, ultratraces	10 ng/g-10 μg/g	2%	factor of 2 with reference samples: 5%
Proton back-scattering (to 2.5 MeV)	surface layer cross-section 1 mm thin film thickness to 3 μm	non-destructive survey analysis; major constituents to admixtures	10^{-3} monolayer	3%	10% with reference samples: 5%

often be necessary to use more than one method.

The multiplicity of methods and the consequent profuse growth of specializations make chemical analysis a labyrinth to the outsider who wishes to have an analysis made, and he will often be at a loss to know to which particular specialist he should take his problem. As a result there is a risk that his problem will be tackled too one-sidedly.

In our laboratories we have solved this difficulty by instituting a weekly meeting of the various specialists responsible for service activities, at which every analytical problem submitted is discussed and a method or combination of methods to solve each problem is recommended. If the problem requires a great deal of work, or is difficult or important, it may often be desirable to set up an *ad hoc* working group consisting of the analytical chemists involved and the 'client'. This group will then discuss the problem thoroughly and decide on the whole analytical procedure, from taking the samples to establishing the final result. A group discussion of the final result ensures that it is properly interpreted and also provides some guarantee that appropriate use will be made of the analytical information.

It has been found, especially in the development of new materials, that much unnecessary work can be avoided and the fastest progress made if there is close cooperation between the preparative and analytical groups right from the start of the investigation.

Service analysis in the research laboratories of an electronics company

The work which an analytical group is called upon to do in laboratories like ours differs very considerably from that required in an establishment such as a factory. Whereas in a factory series of routine analyses are frequently required, this is not the case in our laboratories. Our analytical problems are characterized by the very wide variety of the substances submitted for analysis, by the very different requirements imposed on the analytical results, and the new and often severe constraints that may apply to the analysis. In this situation it is clear that if the analytical services are to function effectively they must be backed up by both specifically oriented and exploratory investigations.

The variety of materials is so wide in the first place because of the very wide selection of materials involved

in solid-state research for an electronics company. They include metals and metallic compounds, semiconductors, glasses, and also ceramics, such as the oxidic magnetic materials.

The widely varying requirements relate to the completeness, reproducibility and accuracy of the information needed. A good example is the request for a total analysis of semiconducting compounds of interest for injection luminescence, such as GaP. Injection luminescence is based on electron transitions between centres formed by very specific trace elements. It is affected by the presence of other trace elements, ultratraces and by a deviation from the stoichiometric composition. Exact knowledge of these three factors is therefore obviously desirable. Another example is the desired reproducibility in the determination of ultratraces of transition metals in quartz glass. It is essential to control the concentration of these traces in the ng/g region in quartz-glass fibres for optical waveguides if the losses are to be kept below 10 dB/km, the upper limit of their practical usefulness. Yet another example is the monitoring of impurities in deionized distilled water, which is used in integrated-circuit technology for rinsing the circuits. In this water 20 ultratraces are determined in the concentration region of 10^{-4}-10^{-1} ng/g.

New constraints encountered in addition to new or more difficult requirements for the analysis include the limited quantity of available material, and unsuitability for analysis of the form in which the material is presented. Particular cases in point are thin films (0.01-1 μm), surface layers (< 0.01 μm) and even adsorbed layers down to parts of an atomic monolayer. In the case of adsorbed atoms an indirect determination, e.g. by means of a gas analysis, may sometimes not be sufficient, and a direct *in situ* determination of the atoms may be desirable; the surface available for analysis may then be as small as a few square microns.

An example of a situation in which the amount of material available is very small is the analysis of thin films of doped yttrium-iron garnets for bubble memories, where the ratio Y/Fe has to be correctly known to an accuracy better than 1% in only a few micrograms of material. Another example is the analysis of photosensitive layers of composite materials for camera tubes. In addition to the major constituents of these layers it is also necessary to know the trace elements and ultratraces they contain. Ultratraces of copper, for example, in concentrations of 10^{13} atoms/cm^3 are enough to cause perceptible differences in material properties.

A strong stimulus to the investigation of surface films and adsorbed layers is provided by modern electronics technology and research. The technological stimulus arises from the trend towards planar devices (integrated circuits, circuit elements and wiring), which require scrupulous attention to surface properties. The stimulus from fundamental research comes from the realization that the surface layer of a body is a separate 'phase', with properties that differ from those inside the body (the bulk properties) and may be influenced by adsorbed atoms. Just as deviations from the ideal geometric structure — vacancies, grain boundaries, dislocations, etc. — partly determine the properties of a solid, so also do the analogous deviations from the homogeneous chemical composition, in particular segregations from a second phase and the occupation of grain boundaries and dislocations by adsorbed atoms. The *in situ* determination of minute quantities of material is obviously important here.

Significant constraints that may occur in monitoring a process are the need for a non-destructive analysis that does not interrupt or interfere with the process, the requirement for automated monitoring, and the requirement that the available time and costs should not exceed a specified maximum. Such constraints will of course severely limit the choice of analytical methods, and in many cases may even make new investigations necessary.

Analytical research

This last section of this article will give an indication of the main lines along which research in the field of inorganic chemical analysis is being undertaken. To begin with, the improvement and extension of existing methods will be discussed. It will again be convenient, as in the previous section, to use the classification of chemical analysis into single-element methods and simultaneous multielement methods, confining the treatment to subjects we are studying at Philips Research Laboratories. The article ends with a few words about longer-term innovations and the principles that can be applied in entirely new methods of analysis in the future.

Single-element determinations

In single-element analysis significant developments are taking place in the applications of microanalysis and ultramicroanalysis: for example the determination of trace and ultratrace elements and stoichiometry determination in very small quantities. Important progress is also being made with methods for the determination of atomic groups and atomic charges. These various developments will now be briefly reviewed.

Stoichiometry determination

Research on semiconducting materials is being extended from binary to ternary and quaternary com-

pounds. These new materials have a variety of technological applications, and are frequently applied in the form of thin films, whose composition and doping profile is essential to the operation. This often entails the analysis of extremely small quantities of material. Whereas macroquantities (10-100 mg) can be used for the stoichiometry determination of GaAsP for a diode laser, only microquantities (1-10 μg) are available for the analysis of photosensitive layers in camera tubes, and ultramicroquantities (μg-ng) for the investigation of composition and doping profiles. The availability of methods for the latter investigations is moreover important for stoichiometry analysis as a whole, including analyses on larger samples, since in stoichiometry determinations an accuracy is sometimes desired and obtained that is meaningless because of the inhomogeneity of a layer sample [1].

Atomic absorption spectrometry (AAS) using furnace techniques seems to be the most promising line of further development for stoichiometry analyses of very small quantities, in particular because a relatively large number of elements can be determined with this method. The method is highly sensitive and has a very low detection limit. In the ng region it is now possible to reduce the random uncertainty of this method to a few per cent, and mechanized sample handling will reduce it still further. The considerable uncertainty of AAS applied to composite samples (see Table I) indicates that further research is required.

Other methods requiring renewed attention for their high sensitivity and the low concentrations at which they can be used are spectrophotometry and coulometry. Further study of reaction kinetics to extend its useful range of applications would also be worth while.

Determination of valency or charge

In the investigation of magnetic materials, both in macro and in microquantities (thin films) it is important to be able to determine the valency or charge of a number of elements together. Chemically this can only be done for one element. The dangers of 'wet chemistry' for valency determinations have already been mentioned in the introduction. It therefore seems important to investigate the possibilities of solid-gas reactions. Thermodynamic (equilibrium and kinetic) investigations are needed with a view to mastering the heat treatment of ultramicroquantities of material and to analysing very accurately the gas atmosphere with which the materials are in equilibrium. For investigations of valency and charge it is also very important to pursue studies of the signals in Auger electron spectroscopy and ESCA that depend on the state of linkage or bonding of the elements.

Determination of atomic groups or radicals

The determination of groups of atoms (such as complex anions or radicals) calls for the development of selective electrodes in addition to the classical chemical methods of titrimetry and gravimetry. Selective electrodes have long been in use for cation and anion determinations, making it possible to perform them very quickly by direct potentiometric methods or by potentiometric titration in a wide range of concentrations (10^{-5}-10^{-1} M), and also in ideal cases with very good reproducibility. Notwithstanding their name, many electrodes are not specific, being sensitive to interferences and susceptible to contamination, so that here again further development is necessary.

Simultaneous multielement determinations

In simultaneous multielement determinations the ideal would be the capability of performing a 'total analysis', using one method to determine all elements qualitatively and quantitatively with great accuracy in a concentration interval of nine decades. Even if a suitably sensitive method existed capable of giving the required reproducibility, it would only be possible to achieve this goal by eliminating or making allowance for matrix and interelement effects. With matrix effects the changes that can take place in the atomic signal as a result of differences in the state of linkage or bonding of an element must be considered, as well as the effect that the atoms as a whole may have on the signal. Interelement effects involve the problem of signal overlapping. We shall now look at some developments likely to take place in the principal methods in the near future.

Atomic emission spectrometry

By means of atomic emission spectrometry using a d.c. arc some 70 elements can be quantitatively determined in an interval of seven decades. The total accuracy obtained by refering to elemental standards can be better than a factor of 3.

Exact calculations of interfering effects are not yet possible, and therefore comparison with reference samples is required for a more accurate determination. The $LiBO_3$ flux method [2] represents a successful attempt to avoid both the matrix effects mentioned above, on the one hand through the use of solutions to get around the bonding problem, and on the other by giving the plasma a chemical alkaline buffer. In determinations of trace elements the high blank value — the trace elements deliberately introduced into the

[1] A more detailed consideration of accuracy is given in the article by E. Bruninx and L. C. Bastings in this issue.
[2] F. J. M. J. Maessen; thesis, Amsterdam 1974.

sample with the $LiBO_3$ — will often be an obstacle to the application of this method. Atomic emission spectrometry of solutions with an r.f. coupled argon plasma by P. J. W. M. Boumans' method [3] yields a smaller random uncertainty and also, because of the greater agreement between samples and elemental standards, provides a more accurate determination than the excitation of solids with the d.c. arc. Here again an alkaline buffer solution can avoid the problem of plasma variations and the associated differences in excitation due to matrix variations.

Without preconcentration the required interval of nine decades does not seem feasible and the simultaneous determination of all the elements remains for the time being an unattained ideal. The principal task of atomic emission spectrometry in the near future will continue to be the survey analysis of major and minor constituents and trace elements.

Mass spectrometry

With spark-source mass spectrometry as many as 85 elements can be determined quantitatively in solids in an interval of nine decades. The accuracy of the quantitative determination, however, is no better than a factor of 3. The reproducibility can be increased by substituting electrical detection for the photographic plate. Improvements in the accuracy must mainly be sought in better control of the ionization process, i.e. the spark mechanism. The main objective of this form of mass spectrometry would seem at present to be to improve the accuracy with which it can provide survey analyses of ultratraces.

X-ray fluorescence analysis

X-ray fluorescence can be used for the quantitative determination of 70 elements in an interval of five to six decades. The reproducibility is very good (0.1-1 %) and a satisfactory accuracy can be attained if reference samples are used. In recent years it has also been possible to calculate the matrix and interelement effects [4] but calibration still offers the greatest certainty. The method can be used for the determination of major constituents right down to ultratraces. The main development in the near future would seem to be an improvement of the resolution or reduction of the background for example for the determination of trace elements and ultratraces. Here again the use of solutions permits better agreement between samples and standards, which improves the accuracy of the determinations although to the detriment of the limit of detection.

Neutron activation analysis

Neutron activation analysis combined with gamma spectrometry can be used for the quantitative determination of about 60 elements. The interval of nine decades is not always achieved without chemical separation. The accuracy of the absolute determination of ultratrace elements is capable of further improvement as soon as better nuclear data are available. For certain elements neutron activation analysis is more sensitive, and for many of them it is more accurate than mass spectrometry.

In situ methods for local analysis

The developments in the analysis of thin films and surface layers using EMA, SIMS, AES, ISS and ESCA (see Table I) will in the long run make accurate determinations possible in small volumes or small areas, and hence of small quantities in general. This can be achieved by increasing the stability of the excitation or ionization and by refined calculation of matrix effects by fundamental-parameter procedures.

The accuracy of both single-element and multi-element determinations is likely to be improved and the analysis time shortened by mechanisation of sample handling. Automatic procedures that shorten the analysis time, which have long been in use for multielement determinations [5], need then no longer be limited to the correction, reduction and processing of data, but can be extended to the whole analysis.

New methods

An important new approach is the development taking place in 'kinetic' analysis, in which the function of the instrument is not to record the end-point of a chemical reaction but to measure a reaction rate. The method can be used for determining the substances actually involved in the reaction, their change in concentration then being measured, and also for determining the concentration of a catalyst, inhibitor or activator affecting a reaction speed. This method shows promise of short analysis times and very low detection limits with simple automatic equipment.

Important new procedural developments are also to be expected in the coupling of inorganic chemical processes with biological processes, since these are highly specific, sensitive and fast. An example is the fixation of enzymes on membranes of ion-selective electrodes to improve selectivity, or the use of bacteria for the rapid determination of toxic elements.

[3] P. W. J. M. Boumans and F. J. de Boer, Spectrochim. Acta **27B**, 391, 1972, and Spectrochim. Acta, in the press.

[4] See the article by M. L. Verheijke and A. W. Witmer in this issue.

[5] See the article by A. W. Witmer, J. A. J. Jansen, G. H. van Gool and G. Brouwer in this issue.

Multielement analysis by optical emission spectrometry — rise or fall of an empire?

P. W. J. M. Boumans

The position of emission spectrometry

Multielement analysis versus single-element analysis

Until about 1965 the *term* multielement analysis was hardly ever used. Nevertheless, multielement analyses were indeed *carried out*, using one of the methods suitable for the purpose: optical emission spectrometry or mass spectrometry. These methods enabled the analyst to determine a large number of chemical elements simultaneously in a single sample.

In optical emission spectrometry (OES) 'simultaneous' determination is possible because the signals that are characteristic of the different elements are present in a single spectrum emitted by the excitation source during the time the sample is exposed to the excitation. This spectrum can be recorded in its entirety on a photographic plate or selectively with the aid of exit slits and photomultipliers with integrating devices. The photographic plate or integrators are read out sequentially, of course, but this is something that can be done in a very short time with the technical facilities now available [1]. If the spectrum emitted by the source does not change with time it can also be sequentially recorded with a single detector. It is then only necessary to vary the wavelength setting of the spectrometer during the measurement. The measurements can however be made under the same conditions for the various elements. This can be taken as a characteristic feature of multielement methods.

In single-element methods it is inherently necessary to carry out the measurements sequentially, because the instrument has to be completely readjusted on changing from one element to another.

In accordance with this definition, wavelength-dispersive X-ray fluorescence spectrometry (XFS) and atomic absorption spectrometry (AAS) should in principle be included with the single-element methods. However, technical refinements to the spectrometer can be made that enable the actual spectrometer to be 'completely readjusted' (see above) via simple external controls, or even automatically. The single-element method then appears very much the same as a multielement method to the analyst. This is the case, for example, with wavelength-dispersive XFS. A multielement character is also suggested here by the fact that XFS is non-destructive, so that only a single sample is necessary.

Present-day atomic absorption spectrometry is clearly a single-element method in character, since the equipment — in view of the market requirements — is kept relatively simple. This is not so with wavelength-dispersive XFS — again because of market requirements.

OES has long held a kind of monopoly position for accurate routine analyses of metals and general analyses of materials of different types and composition — mainly nonconducting materials in powder form, but also metal samples and solutions. In the mid-1950s this situation began to change, because of the advent of XFS, AAS and neutron activation analysis. AAS was tailored more specifically to the needs of wet-chemical analysts, who tended to regard single-element analysis as a basic necessity. Analysts of this persuasion did not therefore take the single-element character of AAS to be in any direct way a negative aspect of the new analytical method: its possibilities were simply compared with what had previously been available, and on these grounds it was regarded as a definite improvement.

AAS also took over a part of the field, however, where OES had previously been applied with only moderate success, in particular for the determination of major and minor constituents in nonconducting materials and in solutions. This led to keen competition, with the contendents bandying terms such as 'single-element analysis' and 'multielement analysis', the latter generally being reinforced with the epithet 'simultaneous'. The keenness of the rivalry has tended to increase in recent years. However, the main interest in analytical chemistry is not in the competition between the proponents and opponents of the various methods, but in considering which of the methods provides the best theoretical and economic solution to a given analytical problem. It therefore seems useful and interesting, within the scope of this article, to fill in the background to some of the developments that have taken place in atomic spectrometry before tackling the main theme — new trends in OES, and in particular the multielement analysis of solutions.

Dr P. W. J. M. Boumans is with Philips Research Laboratories, Eindhoven.

[1] Further information on the automatic read-out of photographically recorded spectra is given in the article by A. W. Witmer *et al.* in this issue, page 322.

Analytical methods based on optical atomic spectrometry

Analytical methods based on optical atomic spectrometry fall into three categories, classified by the nature of the process that provides the data for the analysis: emission, absorption and fluorescence. This subject has been dealt with in a previous article [2].

Strictly speaking, however, the nature of the process from which the information for the analysis is derived is less important than the technical and economic potential of the process and its limitations in analytical chemistry. A classification of atomic-spectrometric methods of analysis should not therefore be based primarily on the distinction between the physical principles employed (emission, absorption and fluorescence) but rather on the various factors that describe the analytical and economic performance of the methods. The main questions listed in *Table I* can then be taken into account [3].

When the variations of the methods of atomic spectrometry are classified with the aid of the table, we again see that there are distinct dividing lines between emission, absorption and fluorescence methods. There is apparently a close connection between the physical principles on which a method is based and the usefulness and economy of that method for the analytical chemist. We shall illustrate this with a comparison of emission and absorption methods.

Optical emission spectrometry versus atomic absorption spectrometry

In principle all chemical elements can be determined both by OES and by AAS. Both could also be used for the simultaneous determination of elements. In fact, however, a fairly sharp dividing line has to be drawn, and OES may be characterized as a method eminently suited for simultaneous multielement analysis, whereas AAS is a typical single-element method of analysis. This distinction is based first and foremost on the differences found in the equipment at present available for OES and AAS [4]; see *fig. 1*. In their turn these differences reflect a development in which market factors have been weighed against the basic potentials of OES and AAS for the purposes of chemical analysis.

As we saw above, OES had a virtual monopoly until 1955 [5], since it was possible to market OES instruments for economic multielement analysis that met the requirements of a particular, although relatively small, group of analysts who wanted to be able to determine simultaneously a large number of elements (anything between 10 and 70) in very different concentrations (μg/g up to tens of per cent). There would have been little point in using AAS with a continuous radiator as the light source — even though under certain conditions it would be possible to obtain in this way an absorption spectrum complementary to the emission spectrum. Because of the requirements imposed on quantities such as the resolution of the spectral equipment, the use of AAS would have provided no instrumental or economic advantages, and therefore AAS as a multielement method was not an attractive alter-

Table I. Checklist of the main questions to be considered when choosing an analytical method.

1	What chemical elements does the method cover?
2	Can more than one element be determined at the same time, or do the elements have to be determined separately one after another? In other words: is the method a multielement or a single-element method?
3	Are solid, liquid or gaseous samples used in the analysis?
4	What is the minimum amount of sample needed for an analysis?
5	What are the detection limits?
6	In which concentration ranges can quantitative determinations be carried out?
7	How great is the chance of systematic errors (inaccuracy), due for example to matrix effects?
8	What is the level of reproducibility, for both short and long term?
9	Are there 'memory effects' or problems from blanks? [*]
10	How much time elapses between taking the sample and obtaining the results?
11	How long does it take to prepare the sample?
12	How long is the equipment or chemist occupied with an analysis?
13	How big is the equipment and how much does it cost?
14	Is the equipment simple to operate, reliable and economic in use?
15	Is the method as a whole very complex, and does an analysis require operators with special skills?
16	Can the analysis be carried out mechanically and automated?

[2] P. W. J. M. Boumans, F. J. de Boer and J. W. de Ruiter, Philips tech. Rev. **33**, 50, 1973.

[3] This point is dealt with in detail in the text of a paper presented in German by P. W. J. M. Boumans: Lücken und Brücken in der Spektrochemie, to be published in Zeitschrift für analytische Chemie.

[4] A comprehensive survey of atomic absorption spectrometry equipment on the market before 1972 has been given in C. Veillon, Handbook of commercial scientific instruments, Vol. 1, Atomic absorption, Dekker, New York 1972. Brief descriptions of commercial instruments will be found in Annual Reports on Analytical Atomic Spectroscopy, Vol. 1, pp. 35-41, 1971, Vol. 2, pp. 38-41, 1972, and Vol. 3, pp. 36-41, 1973, published by The Society for Analytical Chemistry, London 1972/1973/1974, and in C. Jongerius and L. de Galan, Chem. Weekbl. **70**, No. 38, L 15, 20 Sept. 1974.

[5] The year 1955 may conveniently be considered here as having been a 'turning point' since it marked the publication of two classic works, by A. Walsh (Spectrochim. Acta **7**, 108, 1955. and by C. T. J. Alkemade and J. M. W. Milatz (Appl. sci) Res. B **4**, 289, 1955). The authors described the potential applications of atomic absorption for chemical analysis when a hollow-cathode lamp is used as primary emission source, i.e. a source radiating very narrow spectrum lines. After 1955 the development of atomic absorption spectrometry was at first slow, as described by A. Walsh in a recent review article (Atomic absorption spectroscopy — stagnant of pregnant?, Anal. Chem. **46**, 698A, 1974). Since 1973, however, when instrument makers also discovered the potential of the new method, the development has been spectacular.

[*] 'Memory effects' refers to the effects of traces of previous samples left in the equipment, and 'blanks' are effects from contaminants introduced in preparing the sample, during the analysis or in preparing reference samples.

Fig. 1. Injecting a sample into the tubular furnace of the Pye-Unicam SP 1900 system for atomic absorption spectrometry. The analytical equipment, set up on the table, consists of four sections. The section on the extreme left contains six spectral lamps, any one of which may be selected as required. The next section contains the control system for feeding an inert gas into the furnace. Then follows the furnace itself, a horizontal graphite tube open at both ends which is located in the light path between the spectral lamps and the extreme right-and section of the equipment. This last section contains a monochromator and a detector. The sample for analysis is dried in the furnace, evaporated and atomized. The detector now determines the fraction of the light from the spectral lamp that is absorbed by the atomized sample. The equipment on the right is for controlling the furnace temperature in accordance with a preset programme. With this equipment it is possible to detect 1 ng of lead in 10 µl of a solution.

native to OES. This is still the case, in spite of the developments that have taken place since 1955: much of the attraction of AAS is its instrumental simplicity, but this simplicity is also related to the single-element nature of the method. Since this limitation is in many cases not considered a drawback, AAS meets a need and has become so popular that it is sometimes claimed that the contest between OES and AAS seems to have been decided in favour of AAS. Let us see what substance there is in this claim.

More often than not the main function of an analytical laboratory is to provide a service, and therefore the costs incurred should be commensurate with the income received for the service. The choice of analytical

methods is thus largely determined by economic factors, i.e. by the answers to the questions (10 to 16) in Table I. The management's decision whether or not to purchase certain instruments ultimately depends on the answers to these questions.

Until the mid-1950s optical atomic spectrometry offered little that could meet the analytical chemists requirements. (The same applied to other existing methods of physical analysis.) The great breakthrough in the development of instruments for the analytical laboratory did not begin until after 1955. At the start of this development the manufacturers of instruments for OES had three main groups of customers. First, there was the metal industry, in particular the steel industry, where there was a need for emission spectrometers that would make it possible to carry out simultaneous multielement analyses of metal samples rapidly and accurately (sometimes better than a relative 1%) and preferably automatically [6]. Next there were the 'chemical' laboratories, including clinical laboratories, where inexpensive and easily operated flame photometers were brought into use for the determination of sodium, potassium and calcium in solutions. And finally, there were the 'spectrochemical' laboratories.

In the spectrochemical laboratories specialists developed analytical methods using equipment that consisted partly of commercial instruments and partly of instruments designed and built by themselves. These methods were used for tackling widely diverse analytical problems. The samples themselves were of many different types, ranging from nonconducting materials such as rock, minerals, ores, vegetable ash, soil samples, pure substances, nuclear fuels and corrosion residues to metals, used lubricants and solutions. There was also great diversity in the number of elements required (up to 70 and more), in the range of concentrations to be covered (from μg/g to tens of per cent) and in the accuracy (i.e. the total uncertainty) and reproducibility (precision) required, which was an order of magnitude in trace analyses, and a few per cent (relative standard deviation) in the determination of major constituents.

The market situation for the first two categories was clearly defined in terms of instrument specifications that would be acceptable and the investment that would be forthcoming. In addition the customers generally confined themselves to providing specifications, but looked mainly to the instrument makers for help in finding a solution to their problems.

The situation for the spectrochemical laboratories was much less clearly defined. This was a kind of 'do-it-yourself' group, typified not only by the complexity of the analytical problems involved but equally by the imagination, the creativity and drive of the spectrochemists. The design and investigation of excitation sources and techniques for introducing samples into these sources [7], and also the interpretation of spectra recorded on photographic plates [8], offered many analysts an opportunity to specialize in this field and contribute towards its development. Understandably this led to the emergence of a bewildering variety of analytical techniques for emission spectrometry, often with highly individual features [9]. Whatever might be thought of this situation, the fact remains that in the twenty years from 1945 to 1965 'do-it-yourself' emission spectrometry made important contributions to the chemical investigation of materials, especially in trace analyses in semiconductors, nuclear fuels and geological samples [10]. Many 'do-it-yourself' methods also finally resulted in routine methods that could be 'transferred' to automated emission spectrometers, so

Table II. The principal excitation sources, optical spectral-analysis systems and detection systems used in emission spectrometry.

Excitation sources	
Flame	propane-air; acetylene-air; acetylene-nitrous oxide; acetylene-oxygen
Continuous d.c. arc	free-burning d.c. arc in air; gas-stabilized (usually with an inert gas, often with the addition of oxygen); stabilized with water-cooled discs; stabilized with a magnetic field; stabilized with high current; plasma jet
Intermittent d.c. arc	
A.C. arc	with or without ignition circuit
Capacitor discharge ('spark')	unipolar or oscillatory in air; in a special atmosphere (usually inert gas); in vacuum; self-igniting (high-voltage spark) or with a separate ignition circuit; with or without separation between the charging and discharging circuits of the capacitor
Glow discharge	
Hollow-cathode discharge	
Inductively coupled h.f. plasma (1-100 MHz)	
Inductively coupled microwave plasma (2450 MHz)	
Capacitively coupled microwave plasma (2450 MHz)	
Laser	
Optical systems	
Filter(s)	
Interferometer	
Monochromator	manually adjustable; programmable (for a limited number of wavelengths); slow or fast scanning (over the whole wavelength range)
Spectograph	(using photographic emulsion as detector)
Spectrometer	(using photoelectric detection) with several exit slits and detectors (polychromator); with several exit slits and one detector; with movable exit slit and detector; without exit slit and using a TV camera tube or a photodiode array as detector
Detection systems	
Eye	
Photographic emulsion	
Photomultiplier(s)	
Photodiode(s)	
Phototransistor(s)	
Photodiode array	
TV camera tube, with or without image intensifier	
Image-dissector tube	

that once again there was a clearly defined market situation. This was the case in the analysis of substances such as used lubricants (monitoring of wear processes), vegetable ash, soil samples and nuclear materials.

The (emission) spectrum analysts in the 'do-it-yourself' sector also turned their attention to the investigation and development of new excitation sources that could be used in areas where AAS would be fundamentally inadequate. This development was partly stimulated by the gradually clearer recognition of both the weak and the strong points of AAS. The weak points were its single-element character, the 'chemical' interelement effects inherent in the use of a combustion flame as an atomizing cell, and the poor detection limits of elements that form compounds incapable of being dissociated. The strong points were the use of samples in the form of solutions, the simplicity and relatively low cost of the equipment, and last but not least the almost ideal cooperation between the makers of AAS instruments, the users and the 'defecting' spectrochemists. This enabled the instrument makers to gear themselves to market requirements and to keep some control over diversification in the areas of instrumentation and techniques.

A fresh challenge to the supporters of emission methods came with the introduction of analytical flame spectroscopy using hot flames (up to 3000 K) [11] and tubular furnaces as atomizing cells [12], and the emergence of flame fluorescence spectrometry [13] as an entirely new analytical method in atomic spectrometry.

In spite of all this, OES lives on. To some extent this is because the 'market' continues to produce new analytical problems to which only OES offers an appropriate solution, and which would (and does) entail makeshift or strained constructions if other methods are used.

Thus the advent and development of AAS have shown that the use of OES in certain areas, for want of better alternatives, has not always been ideal. The future will have to show whether the present tendency towards the 'stacking' of AA spectrometers to carry out a multielement analysis is nothing more than history repeating itself.

In view of the misunderstandings in this field it seemed to us useful to review a number of relevant aspects of OES, with particular reference to three important questions. The first concerns the extent to which the correctness of the results of an analysis may be adversely affected by the physical state and chemical composition of the sample. The next question concerns the methods of solving this problem, for example by making use of techniques based on solutions or melts, and the appropriate choice of excitation source or excitation conditions. Thirdly we shall consider why OES is particularly suitable for multielement analysis.

Chemical analysis by emission spectrometry

The instruments

The analytical instruments used for emission spectrometry may be divided into those that produce the excitation, those used for the spectral analysis of the emitted light, and those used for its detection and for recording and measuring the spectral data. *Table II* gives a summary of the principal excitation sources, optical spectral-analysis systems and detection systems.

---

[6] Detailed information on the use of automatic emission spectrometers for production control in the iron and steel industry is given in an article by H. van den Berge, N. Kemp and A. W. Witmer: Automatic spectrochemical analysis as a production process control tool in the iron and steel industry, Philips Serving Science and Industry, Oct. 1974.

[7] For a review of excitation sources and research on them, see:
P. W. J. M. Boumans, Excitation of spectra, chapter 6 in: Analytical emission spectroscopy (ed. E. L. Grove), part II, Dekker, New York 1972;
V. A. Fassel, Electrical 'flame' spectroscopy, 16th Coll. Spectrosc. Intern., Heidelberg 1971, Plenary lectures and reports, p. 63, Hilger, London 1972:
and the literature mentioned in notes [3] and [22].

[8] Some idea of the problems involved in the evaluation of photographically recorded spectra is given by P. W. J. M. Boumans in: Enkele fundamentele aspecten van de spectrochemische analyse met de gelijkstroomboog, (Some fundamental aspects of spectrochemical analysis using a d.c. arc), thesis, Amsterdam 1961. See also the article by A. W. Witmer *et al.* in this issue, p. 322.

[9] A revealing example of this is given in a review article by N. Kemp, Experimenteel beproefde werkwijzen, bruikbaar in een algemene spektrochemische analysemethode (Experimentally tested procedures usable in a general spectrochemical analytical method), Belg. chem. Ind. **23**, 615, 1958.

[10] See:
N. W. H. Addink, DC arc analysis, Philips Technical Library, Macmillan, London 1971;
J. Kroonen and D. Vader, Line interference in emission spectrographic analysis. A general emission spectrographic method including sensitivities of analytical lines and interfering lines, Elsevier, Amsterdam 1963; and the article of note [9].

[11] M. D. Amos and J. B. Willis, Spectrochim. Acta **22**, 1325 and 2128, 1966. See also:
J. B. Willis, Atomic absorption spectrometry, chapter 10 in: Analytical flame spectroscopy, Selected topics (ed. R. Mavrodineanu), Philips Technical Library, Macmillan, London 1970.

[12] B. V. L'vov, Spectrochim. Acta **24B**, 53, 1969.
B. V. L'vov, Atomic absorption spectrochemical analysis, Hilger, London 1970.
H. Massmann, Spectrochim. Acta **23B**, 215, 1968.
H. Massmann, 16th Coll. Spectrosc. Intern., Heidelberg 1971, Plenary lectures and reports, p. 285.
G. F. Kirkbright, Analyst **96**, 609, 1971.
F. J. M. J. Maessen and F. D. Posma, Anal. Chem. **46**, 1439, 1974.

[13] J. D. Winefordner and R. Smith, Atomic fluorescence flame spectrometry, chapter 11 in: Analytical flame spectroscopy, Selected topics (ed. R. Mavrodineanu), Philips Technical Library, Macmillan, London 1970.
R. Smith, Flame fluorescence spectrometry, chapter 4 in: Spectrochemical methods of analysis (ed. J. D. Winefordner), Wiley-Interscience, New York 1971.
R. F. Browner, Atomic-fluorescence spectrometry as an analytical technique: a critical review, Analyst **99**, 617-644, 1974.

Various electronic devices are of course used for measurements and processing data, such as d.c. amplifiers, phase-sensitive amplifiers, photon counters, integrators, digital and analog meters, computers with or without feedback, and manually operated, semiautomatic or fully automatic systems for reading out photographically recorded spectra [1].

The various alternatives presented in Table II give nothing like a complete picture of the possible variations. In the case of the 'spark' for instance, a further division can be based on the manner in which the sparking process is electronically controlled (number of sparks per unit time, duration of the separate sparks, the regularity with which the sparks follow each other, the atmosphere in which the discharge takes place, and components are put together to form a complete analytical system are specific instruments produced, such as emission spectrometers for metal analysis [16] (*fig. 2*) or flame photometers for the determination of sodium, potassium, calcium etc. [17].

The situation for excitation sources is quite different; here we have a virtually specific field of *analytical* spectrometry. This may perhaps seem paradoxical, because 'excitation' in the strictest sense, which implies raising atoms from a ground state to higher energy levels, is certainly not confined to the domain of chemical analysis. For the analytical chemist, however, the terms 'excitation' and 'excitation source' have a much wider connotation than for the physicist, as will be explained below.

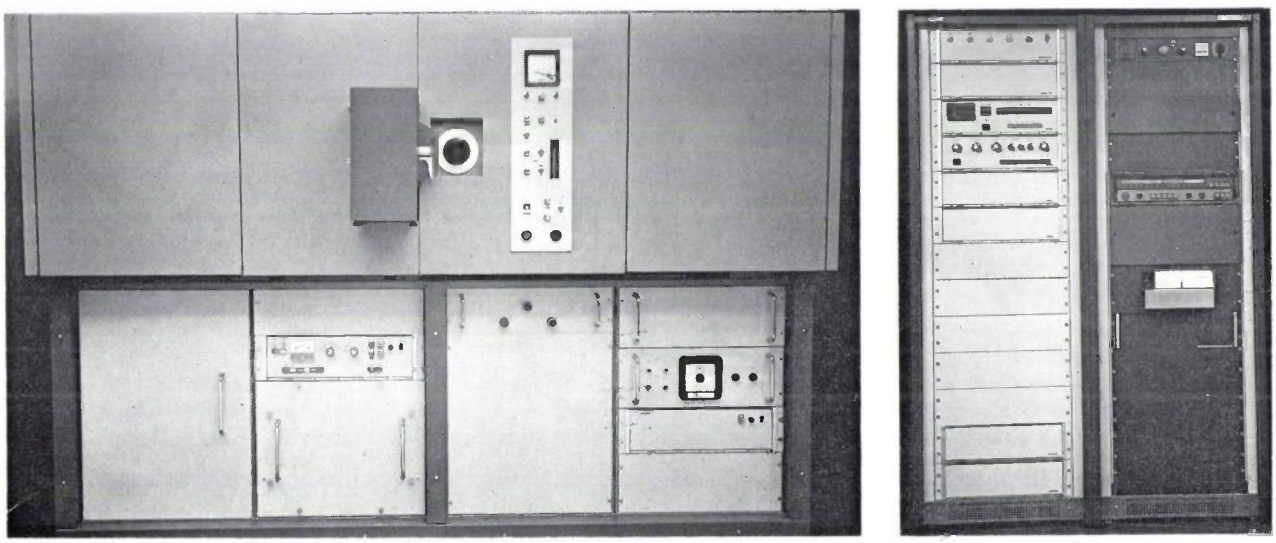


Fig. 2. Vacuum emission spectrometer for automatic multielement analysis of steel. This instrument, made by MBLE in Brussels, enables concentrations of 10 to 20 elements to be determined in less than 10 seconds. The photograph on the left shows the actual spectrometer (above) with the operating equipment. Excitation is by sparking from a disc-shaped sample. The spectrum is detected with photomultipliers. More than 80 of these can be fitted, so that the equipment can be used for analysing many types of steel. The photograph on the right shows the automatic measurement panel (*left*) and a Philips P855 computer, completely programmed for the analysis (*right*). The results of the analysis are presented in digital form on a display panel.

so on) [14]. For the optical systems further differentiations can also be made on the basis of the nature of the dispersion medium (prism, grating, echelle, echelle with prism), the imaging system (Ebert, Ebert-Fastie, Czerny-Turner, Paschen Runge, Wadsworth, Littrow) and the dimensions of the spectral equipment (focal lengths of lenses or mirrors, aperture ratio etc.) [15].

The applications of components in the categories of spectral analysis, detection, measurement and spectral-data processing are not generally limited to emission spectrometry for chemical analysis, but are also found to be in other areas of spectrometry. Only when the

Introducing the sample into the excitation source

The source of excitation is the crux of every analytical technique in emission spectrometry. The source is required to do three things: to evaporate the sample for analysis, atomize the vapour and excite the free atoms and ions formed. In nearly all emission methods these processes occur simultaneously and are initiated in the same way, for example thermal heating. Although incidental attempts have been made to separate the evaporation and atomization from the excitation in both space and time, the techniques developed have not made much headway as yet [18]. The fact has had

to be recognized that these three processes: vaporization, atomization and excitation, are very difficult to separate. This has given rise to the concept of 'sample excitation' [19] [20], a term that covers a complex variety of processes that the analyst must try to control so as to cause samples to emit spectra that can be unambiguously interpreted both qualitatively and quantitatively.

In this context 'unambiguous interpretation' implies that

— the spectra must be reproducible;

— the spectra of both analysis and reference samples must be quantitatively comparable in regard to the intensities of the spectral lines used for the anaysis, to avoid systematic errors in the results;

— spectral interference, 'blanks' and background radiation must not give rise to errors in the results.

Experience in the development of excitation sources has shown that vaporization and atomization give the most problems. Once these processes are properly under control, reproducible excitation and emission follow practically as a matter of course. The investigation and development of a source is therefore inevitably bound up with the investigation and development of techniques of introducing the sample into the plasma of the source. (The plasma is the hot partly ionized gaseous mass in which atomization and excitation take place.)

The method of introducing a sample into a plasma is mainly determined by the nature, state and shape of the sample. Fundamentally, four categories can be distinguished: conductors in the form of a piece of metal, conductors in powdered form, nonconducting solids, and solutions. Gaseous samples are disregarded in this context, since the spectrometric analysis of gases is different in various respects and requires separate treatment [21].

Nonconducting solids and conducting powders are usually investigated with the aid of a d.c. arc and are introduced into the arc by vaporization from a cavity in a graphite electrode. Interrupted spark-like discharges are used for the investigation of solid metal samples. The sample is first recast into a shape in which it can be used as the electrode. At each spark a fraction of the material is released from the electrode and injected into the spark plasma [22]. Solutions are frequently nebulized with a pneumatic or ultrasonic nebulizer and then sprayed into the plasma as a wet or dried aerosol. In certain conditions this can be done in a very uniform and reproducible way, as when a pneumatic nebulizer is used in conjunction with a combustion flame. This combination is a very effective tool in AAS, the flame then of course serving purely as an atomization cell and not as a source of excitation.

Combustion flames have also been developed as an excitation source, but their field of use is limited, as we shall presently see. In particular, flames of this type are not suitable for multielement analysis. The only really effective way of performing multielement analyses of solutions is to use a high-temperature excitation source (> 5000 K). In general, however, aerosol injection presents more problems with high-temperature excitation sources (> 5000 K) than with flames of up to 3000 K. This is due to the high viscosity of such plasmas [23]. In one of the more recent excitation sources, the inductively coupled h.f. plasma (ICP), this difficulty is very effectively avoided by making the argon plasma toroidal in shape, so that the aerosol can be introduced uniformly and reproducibly through the 'tunnel' in the

---

[14] Further particulars of recent developments will be found in Preprints 16th Coll. Spectrosc. Intern., Heidelberg 1971, part I, p. 122 (R. H. Tyas and E. D. Fowler) and 127 (P. Höller, Chr. Thoma and U. Brost), and part II, p. 363 (K. Hollenberg and J. van Calker) and 369 (W. W. Schroeder and A. Strasheim), and in Preprints 17th Coll. Spectrosc. Intern., Florence 1973, part III, p. 112 (J. Schmitz).

[15] For further particulars see:
H. W. Faust, Prism systems, spectrographs, and spectrometers, and R. M. Barnes and R. F. Jarrell, Gratings and grating instruments, chapters 3 and 4 in: Analytical emission spectroscopy (ed. E. L. Grove), part I, Dekker, New York 1971;
J. F. James and R. S. Sternberg, The design of optical spectrometers, Chapman and Hall, London 1969;
P. Bousquet, Spectroscopy and its instrumentation, Hilger, London 1971;
W. Müller-Herget, Some considerations on optical design and selection of spectroscopic instruments, chapter 3 in: Analytical flame spectroscopy; Selected topics (ed. R. Mavrodinenau), Philips Technical Library, Macmillan, London 1970.

[16] A concise review of commercial instruments is given in Annual Reports on Analytical Atomic Spectroscopy, Vol. 1, pp. 29-35, 1971, Vol. 2, pp. 31-37, 1972 and Vol. 3, pp. 30-36, 1973, The Society for Analytical Chemistry, London 1972/1973/1974.

[17] In many cases AAS equipment can also be used for emission measurements on appropriate elements. Reference is made in [4] to sources with data on AAS equipment. The subject of flame photometry is treated in detail in Analytical flame spectroscopy (ed. R. Mavrodineanu), Philips Technical Library, Macmillan, London 1970.

[18] An exception is the laser microprobe, which is used for 'local analyses' of various kinds in geology, for example, and in forensic work. In this analytical method local evaporation of the material is brought about by a high-energy laser pulse, and excitation is produced by means of a subsequent spark discharge between carbon electrodes. For further information see H. Moenke and L. Moenke-Blankenburg, Laser microspectrochemical analysis, Hilger, London 1973.

[19] P. W. J. M. Boumans, Theory of spectrochemical excitation, Hilger & Watts/Plenum Press, London/New York 1966.

[20] See the first article of note [7].

[21] O. P. Bochkova and E. Ya. Shreyder, Spectroscopic analysis of gas mixtures, Academic Press, New York 1965.

[22] Some less conventional methods of feeding solid samples into the excitation zone (including sputtering in a glow discharge and evaporation by laser pulses) are discussed in a review article by K. Laqua, Spektrochemische Lichtquellen, Ein Fortschrittsbericht (Spectrochemical light sources, A progress report), Preprint 17th Coll. Spectrosc. Intern., Florence 1973.

[23] E. Kranz, Proc. 15th Coll. Spectrosc. Intern., Madrid 1969, part 4, p. 95, publ. Ibérica, Tarragona, 34-Madrid-7, 1971.
E. Kranz, Spectrochim. Acta **27B**, 327, 1972.

toroid [24–27]. Various other new excitation sources, including a capacitively coupled microwave plasma [28] [29] and some types of d.c. plasmas [3,30], also make it possible to inject aerosols uniformly and reproducibly into the plasma. Judging from the detection limits reached with these methods, it seems fair to say that the ICP provides the most efficient method of injection. This has recently been shown in a series of unambiguous tests [31].

Homogeneous metal specimens can also be introduced into the discharge region uniformly and reproducibly. The correct choice of the electrical parameters of the spark-like discharge is important here. The effect of fluctuations in the evaporation process can also be significantly reduced by applying a reference element.

Many more difficulties are encountered when powdered specimens are investigated with a d.c. arc discharge. The vaporization process is not so easy to control, both for the constancy of the process as a function of time and for the extent to which the compositions of the vapour is defined by the chemical — i.e. elementary — composition of the solid sample. The chemical composition of the vapour depends not only on the chemical composition of the sample, but may also be affected by its physical properties, such as its structure and grain size.

Solid versus solution

Matrix and interelement effects

In quantitative determinations carried out by a physical analytical method it is generally necessary to use a calibration curve, obtained from measurements on a number of samples of known composition. To obtain correct results, however, it is not enough to know the exact chemical composition of these reference samples. Generally speaking, materials with the same content of a particular element only give an identical signal when they are closely alike both chemically and physically. The requirement of identity implies that 'natural' samples cannot usually be analysed with the aid of a calibration curve obtained from 'synthetic' reference samples. In some cases difficulties even arise when the samples are of the same type but have had a different history, such as steel samples that have undergone different mechanical or heat treatments.

The many and complex problems that the analyst encounters here are covered by the terms 'matrix effects' or 'interelement interference'.

As an example of the effect that the physical properties of a material can have on the composition of the vapour, let us consider the case in which the d.c. arc between carbon or graphite electrodes acts as the source of excitation. The composition of the arc plasma is determined by the atmosphere in which the discharge takes place and the composition of the sample, which is vaporized from a cavity in the lower electrode. The vaporization is generally accompanied by fractional distillation, which means that the composition of the arc gas varies from one moment to another. The vaporization of the material is also usually irregular, and therefore the vaporization curve is not readily reproducible. In an analysis the analysis and reference samples are consequently completely evaporated, the assumption being that there is a unique relation between the time-integrated intensities of spectrum lines and the masses m_A, m_B, ... m_Z, of the elements A, B, ... Z present. A fact that seems remarkable at first sight is that the masses m_A, m_B, ... m_Z do not become fully 'available' for excitation in the arc plasma; only certain fractions a_A, a_B, ... a_Z are 'effective' in this respect, while the fractions $1-a_A$, $1-a_B$, ... $1-a_C$ are 'lost'.

These losses can be attributed to various causes, such as incomplete dissociation of grains that are 'sucked' into the arc, vapour transport through the outer zones of the arc, the spattering of crystallites causing the scattering of material, and the diffusion of constituents of the sample into the wall or base of the electrode, which can give rise to the formation of stable carbides. The course taken by these processes for each of the elements A, B, ... Z depends not only on the specific properties of the elements, such as the volatility of the metal or oxide, but also on the presence of other elements, and in particular on the bonding between the various constituents. If the atoms of an element form part of a crystal lattice, the proportion of the atoms that reach the excitation zone may be entirely different from the proportion reaching it when the atoms are present in the form of a 'separate' finely distributed oxide on the surface of a crystal, even though the material has been identically pulverized in both cases. (Almost complete elimination of the differences is only possible after an extremely high degree of pulverization).

Considering only the grain sizes used in practical analyses, we can say that not only can there be differences between the fractions a_A, a_B, ... a_Z that characterise the interelement behaviour in the same type of material, but that differences may also occur between the fractions a_A, a_B, ... a_Z that become available for excitation in 'natural' samples and the fractions a_A' a_B', ... a_Z' that are effective in 'synthetic' reference samples [32].

The difficulties of matrix effects can sometimes be avoided by using accurately analysed reference samples that are identical in all respects with the samples for analysis or by adopting methods of excitation in which these effects are minimized. But since these ways around the problem are not always available other ways and means have to be found. These include the use of analysis and reference samples fused with lithium metaborate, giving rise to isoformation, or the use of solutions only.

There is no denying that both these methods have some drawbacks. Both fusion and the use of solutions involve the addition of chemicals that can cause contamination. This could be disastrous, especially if traces of commonly encountered elements (such as silicon, aluminium, magnesium, sodium, potassium or calcium) have to be determined in materials of very high purity. The use of fusion or solution techniques also introduces additional manipulations, which may equally well lead to contamination. In addition, these

manipulations may be time-consuming and cumbersome, especially in the analysis of substances that are not readily soluble. In fusion (and in reduction to ash followed by dissolving out) there is a danger of loss of some elements by evaporation.

Without making a generalization, it is true to say that these disadvantages may often be more than compensated by three important advantages. The first is the elimination of matrix effects connected with the structure of solid samples, which makes it simpler to prepare reference samples that lead to correct analytical results. The second advantage is that, because of the homogenization of the samples, there is a very much greater chance that the fraction of the sample that delivers the analysis signal will be identical with the rest of the sample, in other words the results would be more 'representative'. Thirdly, there is the advantage that, since the injection of solutions into an excitation source can be better controlled than the introduction of solids, there is less spread in the analysis signals.

OES again versus AAS; new excitation sources for multielement analysis of solutions

In connection with what has been said above it is interesting to return to the confrontation between OES and AAS. It has already been pointed out that the attraction of AAS lies to a great extent in its instrumental simplicity. This simplicity is partly due to the use of a hollow-cathode lamp as the primary radiation source — so that a small monochromator suffices — and partly to the use of solutions. AAS as originally conceived, in which the solutions were injected as an aerosol into a flame, offered in addition to the advantage of simplicity and easily operated instruments the fundamental advantage that certain matrix effects occurring in solids but not in solutions could be eliminated. Because of this it was much easier to obtain *correct* analytical results with AAS than with most OES techniques. It is therefore not surprising that OES at first lost some ground to AAS, especially in the accurate determination of major and minor constituents.

The progressive development of new sources of excitation, such as gas-stabilized arcs, d.c. plasma jets [3,30], microwave plasmas [28] [29] and ICPs [24–27, 33–36], which enable simultaneous multielement analyses of solutions to be carried out by means of OES, give reason to expect that OES will gradually win back the ground it has lost to AAS where true multielement analyses are required, since these are in fact not possible with AAS, however simple and accurate this method may be.

Another point in favour of OES is that some of the new excitation techniques mentioned above (including the ICP) are not subject to the possible drawbacks of solution methods: the dilution caused in dissolving a solid, which can adversely affect the detection limits. The effect of the dilution can be compensated or outweighed by the efficient injection of the sample into the source and the creation of favourable conditions for excitation in the source itself. This implies that the *relative* detection limits (i.e. the lowest detectable *concentrations* of the various elements) are not only comparable with or even better than those obtained with solid-state techniques (using an arc), but also compare favourably with the detection limits achieved in AAS when the atomization is performed with a flame [37]; see *Table III*. As already mentioned, this is partly attributable to the toroidal shape in which an ICP can be produced [24–27].

There also seem to be good prospects for the *absolute* detection limits, i.e. the smallest detectable *masses* of

[24] See the article by V. A. Fassel, mentioned in note [7].
[25] See the article mentioned in note [2], and also P. W. J. M. Boumans and F. J. de Boer, Spectrochim. Acta **27B**, 391, 1972.
[26] G. W. Dickinson and V. A. Fassel, Anal. Chem. **41**, 1021, 1969.
[27] S. Greenfield, I. Ll.W. Jones and C. T. Berry, Plasma light source for spectroscopic investigation, U.S. Patent No. 3.467.471, 16 Sept. 1969.
[28] W. Kessler, Glastechn. Ber. **44**, 479, 1971.
F. Gebhardt and H. Horn, Glastechn. Ber. **44**, 483, 1971.
M. Sermin, Analusis **2**, 186, 1973.
[29] J. Dahmen and H. Hölzel, 10. Spektrometertagung, Kurzreferate, p. 42, The Hague 1974.
[30] W. G. Elliott, Amer. Lab. **3**, No. 8, p. 45, 1972.
[31] P. W. J. M. Boumans, F.-J. Dahmen, F. J. de Boer, H. Hölzel and A. Meier, Spectrochim. Acta, Part B, to be published; S. J. Baker, Spectrochim. Acta, Part B, to be published.
[32] For a detailed treatment of the problems discussed here and results of experimental work on them, see:
F. J. M. J. Maessen and P. W. J. M. Boumans, Spectrochim. Acta **23B**, 739, 1968;
P. W. J. M. Boumans and F. J. M. J. Maessen, Spectrochim. Acta **24B**, 585 and 611, 1969;
F. J. M. J. Maessen, Enkele aspecten van de spectrochemische sporenanalyse met de gelijkstroomboog (Some aspects of spectrochemical trace analysis with a d.c. arc), thesis, Amsterdam 1974.
R. Rautschke, G. Amelung, N. Nada, P. W. J. M. Boumans and F. J. M. J. Maessen, Spectrochim. Acta B, in press.
F. J. M. J. Maessen, P. W. J. M. Boumans and J. Elgersma, Spectrochim, Acta B, to be published.
[33] R. H. Scott, V. A. Fassel, R. N. Kniseley and D. E. Nixon, Anal. Chem. **46**, 75, 1974.
D. E. Nixon, V. A. Fassel and R. N. Kniseley, Anal. Chem. **46**, 210, 1974.
[34] S. Greenfield, I. Ll. Jones, H. McD. McGeachin and P. B. Smith, Anal. chim. Acta, in press.
[35] J.-C. Souilliart and J.-P. Robin, Analusis **1**, 427, 1972.
[36] P. W. J. M. Boumans and F. J. de Boer, Spectrochim. Acta B, in press.
[37] In 'compromise conditions for multielement analysis' the detection limits for most elements in an h.f. plasma lie between 0.01 and 1 ng/ml in solutions with a matrix. Relating these detection limits to the dissolved matrix we arrive at 0.001-0.1 μg/g (= ppm) for 1 % solutions, and to 0.1-10 μg/g for 0.1 % solutions (see Table III). Further particulars will be found in the articles of notes [33-36], and in:
P. W. J. M. Boumans and F. J. de Boer, An assessment of the inductively-coupled high-frequency plasma for simultaneous multi-element analysis, Proc. Anal. Div. Chem. Soc., in press.
V. A. Fassel and R. N. Kniseley, Inductively coupled plasma-optical emission spectroscopy, Anal. Chem. **46**, 1110A, 1974.

Table III. Detection limits in aqueous solutions [a] as achieved in the ICP after Boumans and De Boer under 'compromise conditions for simultaneous multielement analysis' [b].

Element	Spectral line (nm) [c]	Spectral order	Detection limit (ng/ml = 10^{-9} g/ml)
Al	I 396.15	1	0.2
As	I 228.81	2 [d]	6 [e]
B	I 249.77	2 [f]	0.1 [g]
Ba	II 455.40	1	0.01
Be	I 234.86	2 [d]	0.003
Ca	II 393.37	1	0.0001
Cd	I 228.80	2 [d]	0.2
Ce	II 418.66	1	0.4
Cu	I 327.40	2 [f]	0.06
Cr	I 357.87	1	0.1
Fe	II 259.94	2 [f]	0.09
Ga	I 417.21	1	0.6
Ge	I 265.12	2 [f]	0.5
La	II 408.67	1	0.1
Li	I 670.78	1	0.02
Mg	II 279.55	2 [f]	0.003
Mn	II 257.61	2 [f]	0.02
Mo	I 317.03	2 [f]	0.5
Na	I 588.99	1	0.02
Nb	II 309.42	2 [f]	0.2
Ni	I 352.45	2 [f]	0.2
P	I 253.56	2 [f]	15
Pb	I 283.31	2 [f]	2
Pd	I 360.96	1	2
Sn	I 303.41	2 [f]	3
Sr	II 407.77	1	0.003
Ti	II 334.94	2 [f]	0.03
V	II 309.31	2 [f]	0.06
W	I 400.87	1	3
Y	II 371.03	1	0.04
Yb	II 369.42	1	0.02
Zn	I 213.86	2 [h]	0.1
Zr	II 343.82	2 [f]	0.06

[a] Results are for 0.05 N HCl solutions. Addition of a matrix may degrade the detection limits by a factor of not more than 1.5 to 2, if spectral interference does not rule out the most sensitive line.
[b] Spectrometric conditions: 1-m Czerny-Turner monochromator; grating 1200 rulings/mm, blaze 400 nm, order 1 or 2; slits 25 μm; photomultiplier EMI 9601 B; lock-in amplifier (400 Hz); digital integrator coupled to the retransmitting potentiometer of the recorder; integration time 15 s.
[c] See note [a] in Table V.
[d] With 230-nm interference filter.
[e] In sulphuric-acid solution.
[f] With UG5 filter (Saale-Glas, Jena).
[g] As $Na_2B_4O_7$ in water.
[h] With 210-nm interference filter.

the various elements, although in this respect the flameless AAS techniques (graphite-furnace techniques) continue to give better detection limits for various elements at least, in particular the relatively volatile ones [38].

Here again it is necessary to emphasize the paramount importance of choosing the right analytical technique, basing the choice on such considerations as the particular elements to be determined, the number of elements to be determined in the same sample, the matrices in which these elements are to be found, the amount of material available for analysis and the amount of time that can be spent on an analysis. 'Furnace techniques' should be regarded primarily as microtechniques for the analysis of samples available only in very small quantities and in which neither the matrix nor the elements to be determined present many complications with regard to vaporization, atomization and chemical reactions with the furnace material (graphite). Mention should be made here of a somewhat curious development. The method of analysing a solid with an arc between graphite electrodes is being abandoned, mainly because the thermochemical behaviour of the sample in the graphite electrode gives rise to complicated matrix effects. The alternative chosen to avoid these complications is a solution technique (AAS with a flame as atomization cell), and this has led to an AAS technique using a graphite furnace as atomization cell, in which the solvent is quickly evaporated and the matrix then evaporated in the presence of graphite, so that the analyst is again confronted with the complications arising from the thermochemical behaviour of solids in a graphite environment. An important difference compared with the starting situation, however, is that the thermochemical conditions, and hence the vaporization and atomization, can be better controlled in the graphite furnace than in the graphite arc. On the other hand, the temperatures attainable in the graphite furnace are appreciably lower than in an arc, which implies that the atomization of compounds that are not readily dissociated can present problems [39].

It is therefore clear that the results of analyses using flameless AAS techniques cannot necessarily be assumed to be correct, although this does not yet seem to be generally realized by the users. The successes of flame AAS have generated a kind of belief that analytical results obtained with AAS are just as infallible as the results of wet-chemical analyses. The main grounds for this belief should not however be supposed to reside in the use of AAS as an analytical technique but in the use of *solutions* in a physical analysis technique. The transfer from the flame to the graphite furnace has simply meant that the problems associated with the analysis of solids have 'sneaked into' AAS by another route, thus destroying one of the characteristic advantages of the solution technique, or at least making it very much less certain.

The excitation source for multielement analysis

The importance of the availability of new sources of excitation that can be used for the analysis of solutions in OES has been mentioned above. 'New' here indicates mainly that it is now possible to inject an aerosol into a hot plasma (> 5000 K) in an efficient and reproducible manner while maintaining the stability of the plasma. In itself this does not of course explain why such a plasma is suitable for multielement analysis. This question, incidentally, is not new, since it arises in connection with all excitation sources for multielement analysis, whether classical or not. The question has only been given a new and topical aspect by the association between the analysis of solutions and OES. For this reason we shall now take a more critical look at it.

Atomization and 'chemical' matrix and interelement effects in a plasma

In a multielement analysis the aim is to cover the maximum number of elements with the minimum amount of interference from matrix and interelement effects. As we have seen, matrix and interelement effects connected with the physical properties of the solid state can be avoided by use of solutions. However, there are other matrix and interelement effects that

stem from processes, including dissociation processes, in the plasma of the excitation source.

These effects, often referred to as 'chemical' interelement effects, may result in the first place from the incomplete dissociation of compounds introduced into or formed in the plasma of the excitation source. The effects are described as 'chemical' because they originate in typically chemical equilibria and reactions in the gas phase, including possible catalytic effects on these reactions [40]. The extent to which these chemical effects appear in atomic spectrometric analyses depends on the temperature in the excitation source, the chemical composition of the sample, the nature and composition of the major constituent gas of the plasma, and the time spent by the sample vapour in the plasma before it reaches the observation zone.

In view of the relatively low temperature of a flame produced by combustion (2200-3300 K) chemical interelement effects are particularly prominent in flames, and are therefore more or less characteristic of all flame techniques in atomic spectrometry. (Interelement effects related to the ionization of molecules and atoms in a hot gas are also frequently included with the 'chemical' interelement effects. It is more convenient, however, to treat these as a distinct category. In the following they will therefore be dealt with separately, the term 'chemical' being reserved for interelement effects connected with the dissociation of molecules.)

Shifts in dissociation equilibria, and the occurrence of reactions involving elements to be determined in an analysis, lead to changes in the degree of atomization of these elements. This is reflected in the magnitude of the signals recorded in atomic spectrometric measurements. Whether or not these signals are obtained via emission or via absorption is irrelevant, since the magnitude of the emission or absorption effect is primarily determined by the number of *free* atoms. The relative population of the energy level from which the transition leading to emission or absorption takes place is a secondary consideration.

Chemical interelement effects occur both in OES and in AAS with flames, and to virtually the same extent. A technique using solutions allows the effects of chemical interelement effects on the results of an analysis to be eliminated, since in many cases reference samples with about the same chemical composition as the analysis samples can easily be prepared. The approximate composition of the analysis samples must of course be known. The required effect can also be reached, however, by adding 'releasing agents' to analysis and reference samples. A releasing agent binds an interfering component in the gas phase and thus removes it from the equilibria in which an element to be determined is involved.

A classical example of this is to be found in the procedure used to eliminate the 'interference' caused by the presence of aluminium in the determination of calcium in a flame. If no precautions are taken, a stable calcium-aluminium-oxygen complex is formed, which binds part of the calcium present so that the degree of atomization of the calcium becomes partly dependent on the quantity of aluminium present in the sample. If the composition of analysis and reference samples differs in this respect, this chemical interelement effect can give rise to systematic errors in the analytical results. The addition of a certain excess quantity of lanthanum or strontium as a releasing agent to analysis and reference samples has the effect of binding the aluminium, thus preventing it from interfering with the atomization of the calcium.

In general, attempts are made to achieve as complete an atomization as possible. This is important not only because chemical interelement effects then play a less significant part, but also because the detection power of the method can then be increased, so that more elements can be involved in the analysis and lower concentrations determined.

To increase the degree of atomization, high-temperature flames were developed for AAS. The most widely used is the acetylene-nitrous-oxide flame (3200 K). This development made it possible to include in the analysis elements that were not accessible with the AAS techniques using the earlier flames of relatively low temperature, such as the propane-air flame (2200 K) or the acetylene-air flame (2500 K).

Meanwhile it was found that these hot flames also offered possible uses as excitation sources for OES. An investigation of 68 elements present in pure aqueous solutions showed that the detection limits of 40 of them when an acetylene-nitrous-oxide flame was used

---

[38] The lowest values reported in the literature for the absolute detection limits of Ag, As, Bi, Cd, Pb, Sb, Se, Sn, Te and Tl lie between 0.3 and 300 pg (picograms, 10^{-12} g) when a furnace technique (AAS) is used for samples of 100 µl. See: New Pye-Unicam flameless accessories, Spectrophotometric Information 73/S1, Philips Nederland B.V., Laboratory Instrumentation Group, and:
The HGA graphite furnace increases the scope of atomic absorption spectroscopy, The Perkin-Elmer Corporation, 1974.
For an ICP the absolute detection limits for these elements lie between 10 and 2000 pg, again with samples of 100 µl. For non-volatile elements, however, the detection limits achieved with the ICP are equal to or better than those obtained with furnace techniques. For example, values of 2 and 3 pg are found for Be and Mn, and even 0.03 pg for Ba. See D. E. Nixon, V. A. Fassel and R. N. Kniseley, Anal. Chem. **46**, 210, 1974.

[39] See the book by B. V. L'vov mentioned in note [12]. See also H. Massmann and S. Güçer, Spectrochim. Acta **29B**, 289, 1974, and F. J. M. J. Maessen and F. D. Posma, Anal. Chem. **46**, 1439, 1974.

[40] E. M. Bulewicz and P. J. Padley, Spectrochim. Acta **28B**, 125, 1973.

in OES were just as low or lower than with AAS [41].

Though it may seem from this that flame emission spectrometry now has a wider useful scope than it had some ten years ago, and however attractive the flame may be from the instrumental and economic points of view, a flame is nevertheless not a suitable excitation source for OES multielement analysis. These require a temperature of more than 5000 K, since a source with a temperature of 3000 K is unsatisfactory on three counts. Firstly, elements whose resonance lines have a wavelength shorter than 350 nm are not sufficiently excited. Secondly, the degree of atomization of elements that form stable compounds is insufficient to reach the required detection limits for these elements. Thirdly, the excitation of other than resonance lines is often inadequate, so that the available number of spectrum lines is too small to make it possible to determine simultaneously those elements that occur in strongly different concentrations in the sample. In addition, even at 3000 K, dissociation equilibria and chemical reactions involving atoms of the samples still play a significant role, so that chemical interelement effects of the type described above have to be taken into account. Finally, the molecules of the flame gases emit a large number of band spectra that can use spectral interference. This difficulty is avoided by using an emission source that is not only very hot but also avoids the use of multiatomic gases, such as the ICP with argon as the working gas [24–27] [33–36].

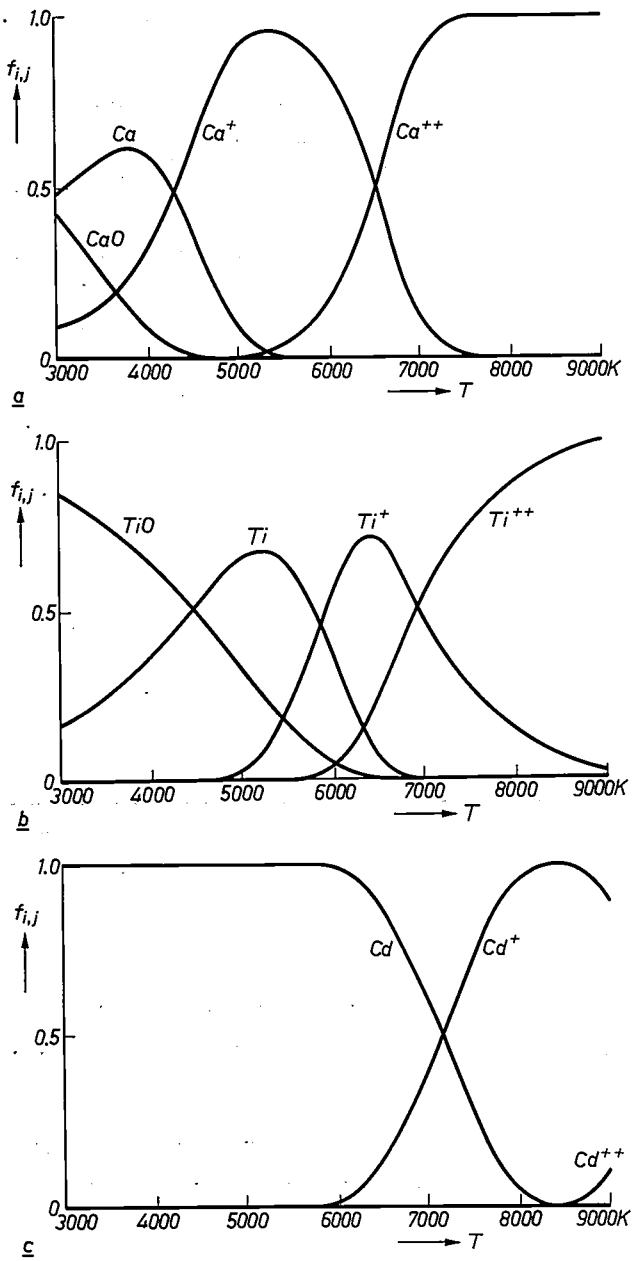


Fig. 3. The fraction $f_{i,j}$ of diatomic oxide molecules, free atoms, singly charged ions and doubly charged ions of Ca, Ti and Cd as a function of the absolute temperature T.

Atomization in an excitation source of temperature > 5000 K

We shall now consider the atomization of the sample in an excitation source in which the temperature is substantially higher than 3000 K [19] [20] [42].

Fig. 3 gives a schematic illustration of the behaviour of three representative elements, Ca, Ti and Cd at temperatures between 3000 and 9000 K in air at 1 atmosphere and for the case where the mass of the element present in the plasma is the same at all temperatures. If we were to start from calcium carbonate, titanium dioxide and cadmium oxide, we should be concerned with CaO molecules, Ca atoms, Ca^+ ions, TiO molecules, Ti atoms and Cd atoms at 3000 K. The relative numbers of particles of the various types mentioned in the figure can be calculated with the aid of equilibrium equations, Dalton's law, the gas laws and the quasi-neutrality equation. Quasi-neutrality implies that the total charge of the positive ions must be compensated by an equivalent number of free electrons.

The difference in behaviour between the three elements as indicated in fig. 3 arises from differences in dissociation energy between the diatomic oxides and differences in ionization energy between the neutral atoms and between the singly charged ions (see *Table IV*). If the temperature of the plasma rises, the relative numbers of particles change: the molecules gradually disappear to give way to neutral atoms and singly charged ions. Increasing ionization leads in turn to a situation where the number or neutral atoms, after an initial rise, starts decreasing. This process is repeated for the singly charged ions as soon as the second ionization stage becomes significant.

If the conditions are chosen such that the temperature in the source lies above 5000 K, it can be seen from fig. 3 that the particles involved are mainly neutral

Table IV. Physical constants determining the behaviour of some elements and their compounds during thermal dissociation and ionization

Compound or kind of particle	Physical constant	Numerical value (°C or eV)
$CaCO_3$	dissociation temperature	825-900
CaO	boiling point	2850
TiO_2	boiling point	2500-3000
CdO	dissociation temperature	900-1000
CaO	dissociation energy	4.2
TiO	dissociation energy	7.2
CdO	dissociation energy	3.8
Ca	ionization energy	6.11
Ti	ionization energy	6.82
Ca	ionization energy	8.99
Ca^+	ionization energy	11.87
Ti^+	ionization energy	13.58
Cd^+	ionization energy	16.91

atoms and ions of successive ionization stages of the elements to be determined in the sample. This means that the atomization is complete, so that on one hand the maximum number of free particles has become available for excitation, which can lead to favourable detection limits, and on the other hand a region has been reached where dissociation equilibria have very little further effect on the available number of free particles, implying that chemical interelement effects are negligible.

Fig. 3 also illustrates that by choosing an excitation source with a temperature higher than 5000 K we have introduced interelement effects of another type, that is to say effects connected with shifts in ionization equilibria. These will be discussed in the next section.

The three curves in fig. 3 relate to the separate systems Ca + air, Ti + air and Cd + air. The curves have not been calculated exactly; they are based only on results of related calculations. As mentioned above, it has been assumed that the mass of each of the three elements present in the plasma remains constant when the temperature changes. Since the pressure has been assumed constant, this implies that the volume of the plasma in the model will increase in proportion to the temperature.

The total number of 'atoms', N_j, of each of the three elements in the plasma is then given by:

$$N_j = \frac{N \times [m_j]_{\text{plasma}}}{G_j} \qquad (1)$$

where N is Avogadro's number, m the mass (assumed constant) and G the atomic weight. The subscript j serves to distinguish between the relevant quantities for the three elements.

The curves in fig. 3 can be represented by a relation of the form

$$N_{i,j} = f_{i,j}(T) N_j, \qquad (2)$$

where the number of particles $N_{i,j}$ of the kind i (molecules, neutral atoms, monovalent or multivalent ions) of an element j is expressed as a fraction $f_{i,j}$ of the total number of 'atoms' N_j. Since the composition and the pressure of the plasma are fixed, $f_{i,j}$ depends here solely on the temperature T. In general, however, $f_{i,j}$ is also function of the composition and pressure of the plasma. Here again there is a source of matrix and interelement effects, as will presently be shown.

Since in this article we are only concerned with the illustration of some general principles, exact calculations have not been used for fig. 3. Such calculations would not in themselves present an insurmountable problem, although even for relatively simple systems they are so complicated that they can only be performed iteratively with a computer. The main difficulty, however, is not so much in carrying out the calculations as in the choice of a model that approximates to a plasma in a true excitation source well enough to enable meaningful conclusions to be drawn. Choosing a model that meets this condition involves taking into consideration the energy balance of the plasma and the spatial distributions of all the parameters concerned. This in turn implies that the particular features of a particular excitation source must be taken into account when formulating the model. Consequently the calculations are highly complicated, especially if they are to be carried out for plasmas of the kind encountered in spectrochemical analysis. Their composition is generally extremely complex because of the complex composition of the samples introduced into the plasma.

Interelement effects in the plasma of an excitation source with a temperature higher than 5000 K

We have seen that in the 'complete atomization' of a sample, as occurs at temperatures above 5000 K, not only neutral atoms are involved but also ions at various stages of ionization. In quantitative analysis this is an important point, because the spectra emitted by atoms and ions are primarily characteristic of the kinds of particle from which they originate (in fig. 3, for example, Ca, Ca^+ or Ca^{++}), and only in the second place characteristic of the element ('calcium'). The intensity of a particular spectrum line is therefore not primarily proportional to the amount of calcium in the plasma but to the number of, say, Ca^+ ions. This number depends not only on the amount of calcium in the plasma but also on parameters that govern the ionization equilibria, in particular the temperature and electron density in the plasma. The numerical values of these parameters are in turn to some extent determined by the chemical composition of the sample introduced into the plasma. If in particular the sample contains constituents that are easily ionized, this can have a marked effect on the temperature and on the electron

[41] E. E. Pickett and S. R. Koirtyohann, Spectrochim. Acta **23B**, 235 and 673, 1968, and **24B**, 325, 1969.
E. E. Pickett and S. R. Koirtyohann, Anal. Chem. **41**, No. 14, 28A, 1969.
G. D. Christian and F. J. Feldman, Appl. Spectroscopy **25**, 660, 1971.
P. W. J. M. Boumans and F. J. de Boer, Spectrochim. Acta **27B**, 391, 1972.

[42] P. W. J. M. Boumans, Fundamental source parameters — measurement, interpretation and application in spectrochemical analysis, Proc. 14th Coll. Spectrosc. Intern., Debrecen 1967, part 1, p. 23, Hilger, London 1968.
P. W. J. M. Boumans, Verfeinerte physikalische Messungen und theoretische Berechnungen zur Deutung von spektrochemischen Beobachtungen und zur Unterstützung von Analysen mit dem Gleichstrombogen (Refined physical measurements and theoretical calculations for the interpretation of spectrochemical observations and in support of analyses with a d.c. arc), 16th Coll. Spectrosc. Intern., Heidelberg 1971, Plenary lectures and reports, p. 268.

density, and consequently on the relative numbers of Ca atoms and Ca^+ and Ca^{++} ions. Because of this the intensity of a spectrum line emitted by, say, Ca^+ ions, in samples containing the same amount of calcium, may differ in the samples that contain different amounts of easily ionized elements. Excitation in a source with a temperature above 5000 K may thus give rise to matrix and interelement effects that are relatively weak or not found at all in flames.

A commonly adopted remedy is a 'spectroscopic buffer', i.e. an easily ionized element, for example an alkali metal, added to samples and standards. The buffer then 'fixes' the conditions (temperature and electron density) in the plasma, and thus at the same time fixes the position of ionization equilibria.

In the case of a d.c. arc this implies that the temperature in the arc plasma is stabilized at a relatively low value by the injection of an alkali metal. This effect, and others associated with it, have been studied in detail, so that at least a good qualitative picture has been obtained of the mechanisms that govern the interelement effects occurring in this source [10] [20] [42]. In the case of other sources, such as the ICP, not even a qualitative interpretation has yet been arrived at [36]. A qualitative study of this source seems inevitably to involve the study of its spatial structure, which is a complicated matter both theoretically and experimentally [43]. Here at all events is an area where further research is needed, not only because of the interest in explaining the effects that occur in a spectrochemical analysis, but also because a balanced exchange between fundamental and applied research would contribute towards the optimum development of these promising excitation sources for the multielement analysis of solutions.

The spectrum of an excitation source with a temperature higher than 5000 K

It will now be shown how it is possible, as already mentioned, to obtain information with an excitation source hotter than 5000 K on which a multielement analysis can be based. To do this we shall consider the dependence of the intensity of the spectrum line on the source temperature in the simple case to which fig. 3 corresponds. One of the decisive factors in this case, the relative concentration $N_{i,j}$ of the particles of one kind, has already been discussed above.

Let us now look at the other important factor, the proportionality factor between the concentration $N_{i,j}$ of particles of the kind i of element j and the intensity $I_{i,j}$ of their characteristic radiation. In a thermal plasma the quantities are related by the equation:

$$I_{i,j}(T) = K \frac{g_q A_{qp} h\nu_{qp}}{Z_{i,j}(T)} e^{-E_q/kT} N_{i,j}. \qquad (3)$$

Here K is an instrumental constant, k Boltzmann's constant, h Planck's constant, E_q the excitation energy of the upper level q of the spectral transition q → p (i.e. the difference in energy between this level and the ground level of the particle of kind i), g_q the statistical weight of the level q, A_{qp} the probability of the transition q → p, ν_{qp} the frequency of the radiation emitted during the transition and $Z_{i,j}$ the partition function (or sum over states) of the particle;

$$Z_{i,j} = \sum_m g_m \exp(-E_m/kT).$$

Table V gives the numerical values of $g_q A_{qp}$, $h\nu$, E_q and $Z_{i,j}$ for a number of spectrum lines of Ca, Ti and Cd. *Fig. 4* shows how the intensities of six of the spectrum lines vary with temperature. It was assumed in calculating the curves that the three elements were present in the samples in the same relative weights and also that the fraction of the sample present in the source was the same for the three elements.

It can be seen in fig. 4 that at temperatures above 5000 K all three elements give spectrum lines of relatively high intensity, and that these lines do not differ much from each other in intensity. At 5500 K, for example, there are three lines: Ca II (396.85 nm), Ti I (498.17 nm) and Cd I (228.80 nm) with a maximum intensity difference amounting to no more than a factor of 3 [44]. At 6000 K there are four lines: Ca II (396.85 nm), Ti I (498.17 nm), Ti II (334.94 nm) and Cd I (228.80 nm), again with a maximum intensity difference of a factor of 3.

A slight difference in intensity is useful in an analysis, especially if the detector has a limited dynamic range as in the case of a photographic plate. It is then easier to determine various elements simultaneously in a single sample.

Table V. Physical constants of some spectrum lines of calcium, titanium and cadmium

Atom or ion [a]	Wavelength (nm)	$g_q A_{qp}$ [b] (10^8/s)	$h\nu_{qp}$ [b] (eV)	E_q [b] (eV)	$Z_{i,j}$ [c] (T = 6000 K)
Ca I	422.67	1.0	2.93	2.93	1.4
Ca II	396.85	0.45	3.12	3.12	2.4
Ca II	315.89	6.9	3.92	7.05	2.4
Ti I	498.17	9.9	2.49	3.34	36
Ti II	334.94	23	3.70	3.75	61
Cd I	228.80	12	5.42	5.42	1.00
Cd I	326.11	0.0090	3.80	3.80	1.00
Cd II	226.50	99	5.47	5.47	2.00

[a] In spectroscopy the neutral atom is indicated by placing the the Roman numeral I after the chemical symbol of the element. For the singly charged ion the numeral II is used, for the doubly charged ion the numeral III, etc.

[b] The values of $g_q A_{qp}$, $h\nu_{qp}$ and E_q have been taken from C. H. Corliss and W. R. Bozman, Experimental transition probabilities for spectrum lines of seventy elements, National Bureau of Standards Monograph 53, published by Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C., 1972.

[c] Values taken from page 94 of the book by P. W. J. M. Boumans [19]. Further information on partition functions will be found in the first article of note [7].

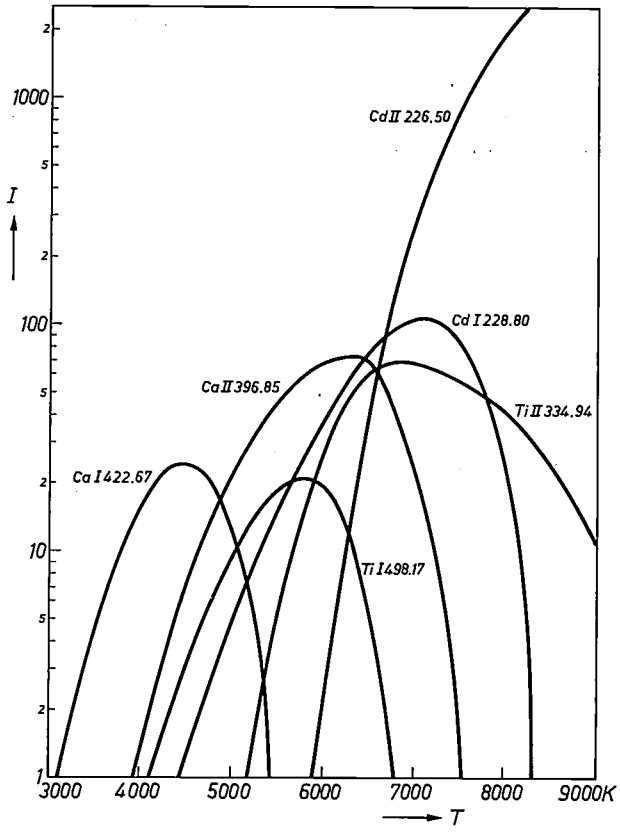


Fig. 4. The intensity I (in arbitrary units) of a number of Ca, Ti and Cd lines as a function of the temperature T of the excitation source. The diagram relates to a sample in which the three elements are present in equal amounts. Between 5500 and 6000 K it is possible, in more than one way, to select a line from each of the spectra such that the intensities of the three lines do not differ by more than a factor of 3.

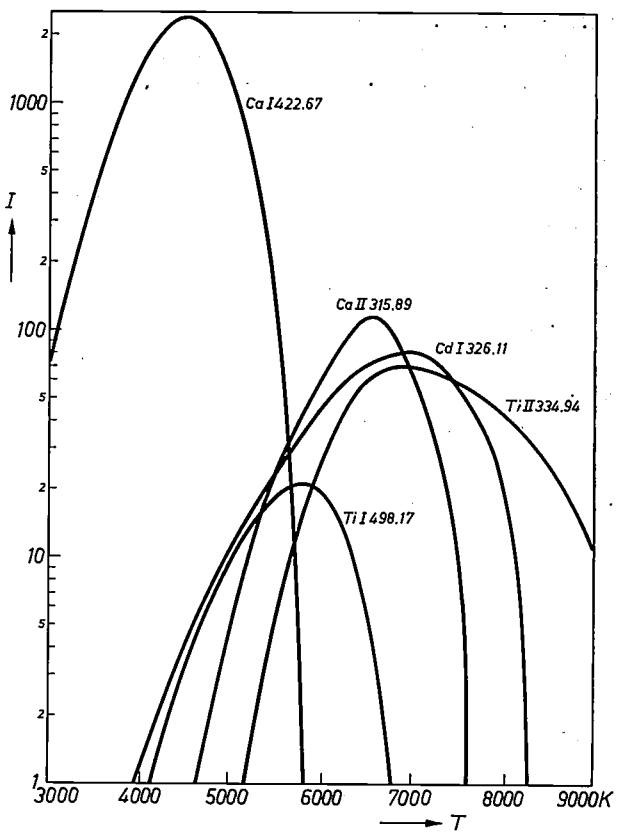


Fig. 5. As fig. 4, but now for the case where the amounts of Ca and Cd are a hundred times greater. Here again, between 5500 and 6000 K it is possible to select from the spectra a Ca, a Cd and a Ti line that differ in intensity by no more than a factor of 3. This makes excitation sources with a temperature above 5000 K ideally suited for multielement analysis with optical emission spectrometry.

Fig. 4 also illustrates that the situation below 5000 K deteriorates rapidly with decreasing temperature. It is now no longer possible to use ion lines to bridge the differences in the physical properties of the elements, which can be done above 5000 K.

If, unlike the case described above, the sample is one in which the elements occur in widely different concentrations instead of in equal quantities, the situation relating to the relative intensities of the spectral lines is not necessarily worse. All that is necessary is to select different spectrum lines. This is because the atomic spectra emitted by sources whose temperature is above 5000 K contain a large number of lines that differ very considerably in intensity. Consequently it is always possible to find lines suitable for the analysis. This is very fortunate since it enables emission spectrometry using a source hotter than 5000 K to offer in principle the full range of facilities for multielement analysis.

An example is given in *fig. 5*. The amount of titanium is the same as in fig. 4, but the amounts of calcium and cadmium are 100 times greater. When the spectrum lines Ca II (315.89 nm) and Cd I (326.11 nm) are chosen for calcium and cadmium (the data for these are given in Table V), a picture is found that is essentially the same as that shown in fig. 4. At 5500 K we have three lines of nearly identical intensity available (Ca II 315.89, Ti I 498.17 and Cd I 326.11 nm) and at 6000 K there are four lines available (Ca II 315.89, Ti I 498.17, Ti II 334.94 and Cd I 326.11 nm) with a maximum intensity difference of a factor of 3.

The emission of spectra containing a large number of lines, although useful for multielement analysis, has the disadvantage that line interference can occur owing to the near coincidence of spectrum lines of different elements. This complication sets a lower limit to the size of the equipment, in particular that of the spectrometer, whose resolution should not be too small. Coincidences of this type can also make the interpretation of analytical results very time-consuming, necessitating automation and the use of a computer.

[43] G. R. Kornblum and L. de Galan, Spectrochim. Acta 29B, 249, 1974.
R. M. Barnes and R. G. Schleicher, Spectrochim. Acta B, in press.
[44] See note [a] in Table V.

Of course, the model on which figures 3, 4 and 5 are based is an oversimplification of the reality. A real excitation source is not homogeneous and cannot be characterized by a single temperature, but only by a temperature distribution. The intensities observed are composed of contributions from regions of different temperatures. Diagrams like those in figs. 3, 4 and 5 should therefore be formulated for real sources with the aid of averages, which results in flatter curves. This does not, however, invalidate the conclusion that emission spectrometry with a source temperature above 5000 K is particularly suitable for multielement analysis. Experience has shown that the model used in the foregoing corresponds reasonably well with the reality and gives a good qualitative picture of the essential features of real sources, so that the conclusions do have a firm foundation. In fact our conclusions here are based on practical experience in emission spectrometry, and we have used a model by way of illustration.

Thermal and non-thermal excitation sources

In the previous section we saw that excitation sources with a temperature of at least 5000 K are particularly useful for multielement analysis by OES. The degree of atomization obtained with it is very high for most elements, there are no serious chemical interelement effects (apart from the effect on the position of the ionization equilibria) and the emitted spectra contain a large number of lines of widely different intensities. This category includes the classical free-burning carbon or graphite arc, plasma jets of various types generated by a d.c. discharge, carbon or graphite arcs stabilized by a gas flow or magnetic field, and very probably one of the more recently developed accessions, the ICP.

The features that make these thermal sources attractive for multielement analysis are also found in various non-thermal sources, such as the spark discharge and inductively or capacitively coupled microwave plasmas. When these sources are used not all aspects of the system can be described by the same temperature. It is frequently found, for example, that there is a temperature T_g, called the gas temperature, which characterizes the average kinetic energy of the heavy particles, and a much higher temperature, T_e, which does this for the free electrons. The temperatures that describe the excitation and ionization, T_{ex} and T_{ion}, are usually comparable with the electron temperature; the dissociation temperature T_{diss} sometimes follows T_g, sometimes T_e. The first situation is found in a spark:

$$T_g \ll T_e \approx T_{diss} \approx T_{ion} \approx T_{ex},$$

the other is probably found in microwave plasmas:

$$T_g \approx T_{diss} < T_e \approx T_{ion} \approx T_{ex}.$$

The magnitude of the numerical differences between the various temperatures in these cases will depend to a great extent on the power dissipated in the plasma.

Emission spectrometry; rise or fall?

From the general tenor of this article, and from what has been said at the end of the last section, it can be deduced that in spite of the emergence of other methods of analysis, emission spectrometry has not been ousted, nor are its days numbered. The title of this article may already have suggested this conclusion. The question was put in this form to enable a number of topical problems to be examined from this particular point of view so as to remove possible misunderstandings.

As explained in the introduction, it is necessary when judging the position and functions of emission spectrometry in analytical chemistry to adopt a differentiated approach, and not to see the situation simply in terms of 'rise or fall?'. Terms such as 'decline', 'consolidation' and 'revival' would be more appropriate to characterize the situation, especially with regard to the optimum matching of the analytical technique to the analytical problem.

'*Decline*' refers to the fact that emission spectrometry has had to give pride of place in certain areas of application (especially in the accurate determination of major and minor constituents in nonconducting materials) to other analytical techniques, in particular to X-ray fluorescence spectrometry and atomic absorption spectrometry. This is a sound development, since the emergence of these methods has added new dimensions to the analyst's potential and has put an end to the use of a technique that offered no effective solution to the analytical problems concerned.

'*Consolidation*' is the appropriate term where emission spectrometry has maintained its position as an indispensable analytical method in other fields, for example for the rapid, direct and accurate analysis of metals, for general overall analysis and for trace analysis. This situation is reflected in the increasing perfection of automatic analytical equipment for the metal industry [6] [16], the development of new excitation sources for the analysis of solids (including various electronically controlled arc and spark sources, gas-stabilized arcs and sparks, magnetically stabilized arcs, glow discharges, hollow-cathode discharges and lasers) [45] and the introduction of equipment that automatically reads out and processes photographically recorded spectra [1].

'*Revival*' characterizes the growing interest in the use of emission spectrometry for the multielement analysis of solutions. The development of excitation sources

with a temperature above 5000 K into which solutions can be injected (for example d.c. plasma jets, the gas-stabilized arc, the inductively coupled h.f. argon plasma and the capacitively coupled microwave plasma) makes it reasonable to expect that emission spectrometry will add a new dimension to the analytical chemist's potential in the not too distant future. This will certainly happen if it proves possible to develop integral analytical instruments that are both technically and economically efficient. This expectation is justified by the mounting demand for multielement analyses and also by the growing realization that the general usefulness of an analytical technique depends not only on its scientific background but also on market factors. The development of systems that meet, both technically and economically, the requirements of analytical chemists who are not specialized in emission spectrometry is therefore an important area in which emission spectrometric activities will concentrate. At the same time, automation and the use of computers will be necessary to bring the evaluation of often complex emission spectra within the range of these analysts.

Summary. This article considers the place occupied by optical emission spectrometry (OES) as a method for multielement analytical chemistry. After initial application in many fields, OES has gradually yielded pride of place to more recent methods in certain applications, yet in others has remained an irreplaceable method. Now the development of excitation sources that can be used in the accurate multielement analysis of solutions has put OES into the limelight again.

The preference that many analytical chemists have for methods suitable for use with solutions is examined in detail, with a discussion of the influence of the physical state and chemical composition of analysis samples on the correctness of the analysis results (matrix and interelement effects). In this discussion it is explained that the use of solutions avoids those matrix effects arising from the specific properties of the solid material, and that a particular group of chemical interelement effects in the gas phase can be eliminated by using an atomization cell at a temperature above 5000 K. It is shown that an atomization cell operating in this range is also an excellent excitation source for multielement analysis by OES. Excitation sources have also successfully been developed that not only run at the correct temperature, but are also designed to permit aerosols of solutions to be reproducibly and efficiently introduced into the plasma. The most promising arrangement here is a toroidal inductively coupled h.f. argon plasma.

[45] See the review article by K. Laqua mentioned in note [22].

Philips tech. Rev. 34, 322-329, 1974, No. 11/12

A system for the automatic analysis of photographically recorded emission spectra

A. W. Witmer, J. A. J. Jansen, G. H. van Gool and G. Brouwer

Introduction

In the analysis of a sample by emission spectrometry atomic spectra in the wavelength range between 150 and 850 nm [1] are used. These spectra are usually generated with an electric arc or spark, but other sources of excitation will also become available in the near future [2]. A qualitative determination of the composition of a sample is made from the positions of the lines in an atomic spectrum expressed in a wavelength scale. Quantitative analyses are based on the intensities of the lines, which are a measure of the concentrations of the constituent elements.

Emission spectrometry is mainly used in two kinds of laboratory: industrial laboratories for the control of particular production processes, and laboratories that provide general analytical services. An industrial laboratory — for example in the steel industry — has to analyse large numbers of samples of only a few types, whose composition may not differ greatly from one sample to another. This usually permits a high degree of automation, since the complete analysis programme can be carried out with relatively few spectrum lines (30 to 40). Direct-reading spectrometers can then be used, usually in conjunction with computers for evaluating the spectral data. Spectrometers of this type have photomultiplier tubes that measure the intensities of light beams of the required wavelengths. These spectrometers work quickly and give reproducible results, but they are not very flexible in the choice of lines for the analysis.

A general analytical laboratory usually has to deal with a large number of different *types* of sample and the composition of the various types may also vary quite considerably. The *numbers* of samples, however, are usually small. In these laboratories emission spectrometry is generally used to give a preliminary overall analysis. This may sometimes provide sufficient information about the composition of a sample. If greater accuracy is required, however, or if smaller quantities of particular elements are to be detected, emission spectrometry then provides the preliminary information for more accurate or more sensitive analytical methods [3]. An example of a general laboratory is the spectrochemical section of the analytical group at Philips Research Laboratories in Eindhoven.

In a general analytical laboratory the preferred detector is a photographic plate, since it is compact and can simultaneously record the light of many spectrum lines, weak and strong. Another advantage of photographic recording over the direct-reading spectrometer is that it is not necessary to specify the analysis lines required when ordering the instrument. The resolution obtainable with a photographic plate is also much higher than can be obtained from a direct-reading system with its relatively wide exit slits.

In the following we shall first briefly discuss the procedure of a conventional analysis in our laboratories, and we shall then show how we have automated the various phases of the analysis.

Conventional analysis

In the method of analysis employed at Philips Research Laboratories [4] we use 400 analysis lines — lines characteristic of a particular element — in the wavelength range from 2450 to 3500 Å [*]. These lines enable 66 elements to be detected and their concentration determined. There are four distinct phases in a conventional analysis.

1) The qualitative analysis. This consists in determining the elements for which analysis lines are present, not including 'overlapping' lines, i.e. lines that coincide with the lines of other elements.

2) Measurement of the blackening (optical density) of the analysis lines by means of a microphotometer [5]. The background blackening is determined for each line: this is the blackening caused at the position of that line by effects other than the radiation to be measured, e.g. weak molecular bands or scattered light. The background blackening is determined by measuring the lowest value close to the line.

3) Conversion of the blackening into light intensities by means of the calibration curve of the photographic emulsion. The line intensities thus found are corrected for background density by subtracting the corresponding intensities.

4) Calculation of the concentrations of the elements

A. W. Witmer, J. A. J. Jansen, G. H. van Gool and Ir G. Brouwer are with Philips Research Laboratories, Eindhoven.

from the intensities of the analysis lines. The intensity I of a line is proportional to the concentration c of the element producing it: $c = qI$. The factor q is usually called the *sensitivity factor*; a stronger line thus has a smaller sensitivity factor.

To calculate the concentration of an element it is necessary to know the sensitivity factors of the various analysis lines of the element. These are found with the aid of a number of reference samples that contain the element in a known concentration. Since the intensity of a line can be affected by the other substances present in a sample (the 'matrix effect'), the sensitivity factors are determined with several reference samples of different composition, corresponding as far as possible to the combinations expected in the samples under analysis.

Considerably experience is necessary in making a qualitative analysis of an unknown sample, and the analysis is therefore usually made by a skilled analytical chemist. On the other hand, the blackening measurements made later for long series of spectrum lines with a manually operated microphotometer, and the calculations of intensities and concentrations, are routine and tedious operations. They also take a great deal of time, so that the results might be adversely affected by human fatigue. It therefore makes sense to eliminate the human factor as far as possible by automating these routine operations, not so much to increase productivity as to achieve better results.

Automatic analysis

In the automatic procedure that we shall now describe the qualitative and quantitative analyses are carried out simultaneously. First, the required information is automatically read from the photographic plate by a specially designed microphotometer and is then input to a computer. This information relates to the sequence of the spectra on the plate (there may be as many as 60 spectra on one plate), the position and blackening of the lines, and also to the background blackening of each line and the location where it was measured.

The information is evaluated with computer programs in which the qualitative and quantitative analyses are combined in a single iterative process, which is necessary because lines from different elements may coincide. Until the composition of the sample is known, it is impossible to know which lines are 'overlaps'. The concentrations are therefore initially calculated from all the observed lines, without taking possible interference into account. This first result is used for determining which lines are significantly overlapped, and the concentrations are then calculated again in a second cycle without using the overlap lines. This result is again used to determine overlap lines, and once again the concentrations are calculated. The procedure is repeated until there is no further change in the calculated concentrations. Comprehensive information is required for these calculations, including tables of the sensitivity factors of the lines of a large number of elements, and tables of interfering lines and molecular bands.

In much the same period as that in which the method described here was developed, work was also under way elsewhere on the automation of spectrum analysis [6]; we have had many fruitful discussions on this subject with our colleagues from Mullard Research Laboratories, Redhill (England), who have developed a similar system for the analysis of mass spectra [7].

The automatic microphotometer

Fig. 1 shows the microphotometer that measures the spectra automatically. Parts of the 'Schnellphotometer' made by Jenoptik of Jena were used in the construction of this instrument. To enable the spectra to be read out at the required speed the selenium cell used in the Jena instrument was replaced by a silicon photodiode, which will operate correctly at frequencies higher than 1 MHz. This is more than sufficient, since the minimum half-width of a spectrum line is about 10 microns, which corresponds to a frequency of at the most 1 kHz at the scanning rate of 1 cm/s.

The photographic plate is mounted on a slide or carriage that can be moved in two directions: these are the longitudinal direction of the spectra, the x-direction, for scanning, and the y-direction for changing from one spectrum to another. The movement in the x-direction must be extremely smooth, and to meet this requirement a system is used that combines air and

[1] P. W. J. M. Boumans, Spektralanalysen: Optische Atomspektroskopie, Techn. Rundschau **63**, No. 37, pp. 49-53, and No. 43, pp. 33-37, 1971.

[2] See P. W. J. M. Boumans, this issue, p. 305.
P. W. J. M. Boumans, F. J. de Boer and J. W. de Ruiter, A stabilized r.f. argon-plasma torch for emission spectroscopy, Philips tech. Rev. **33**, 50-59, 1973 (No. 2).

[3] M. L. Verheijke, this issue, p. 330.
M. L. Verheijke and A. W. Witmer, this issue, p. 339.
J. B. Clegg and E. J. Millett, this issue, p. 344.
E. Bruninx and L. C. Bastings, this issue, p. 350.

[4] N. W. H. Addink, DC arc analysis, Macmillan, London 1971.

[5] N. H. Nachtrieb, Principles and practice of spectrochemical analysis, McGraw-Hill, New York 1950, chapters 4 and 5.

[6] B. L. Taylor and F. T. Birks, Analyst **96**, 753, 1971, and **97**, 681, 1972.
F. G. Walthall, J. Res. U.S. geol. Survey **2**, 61, 1974 (No. 1).

[7] E. J. Millett, J. A. Morice and J. B. Clegg, Int. J. Mass Spectrom. Ion Phys. **13**, 1, 1971.

[*] *Editor's note.* Although the nm is now recommended as the unit of wavelength, we have used the ångström in this article, since it is more generally used in this specialized field and because the wavelengths in the computer programs employed here are all expressed in ångströms.

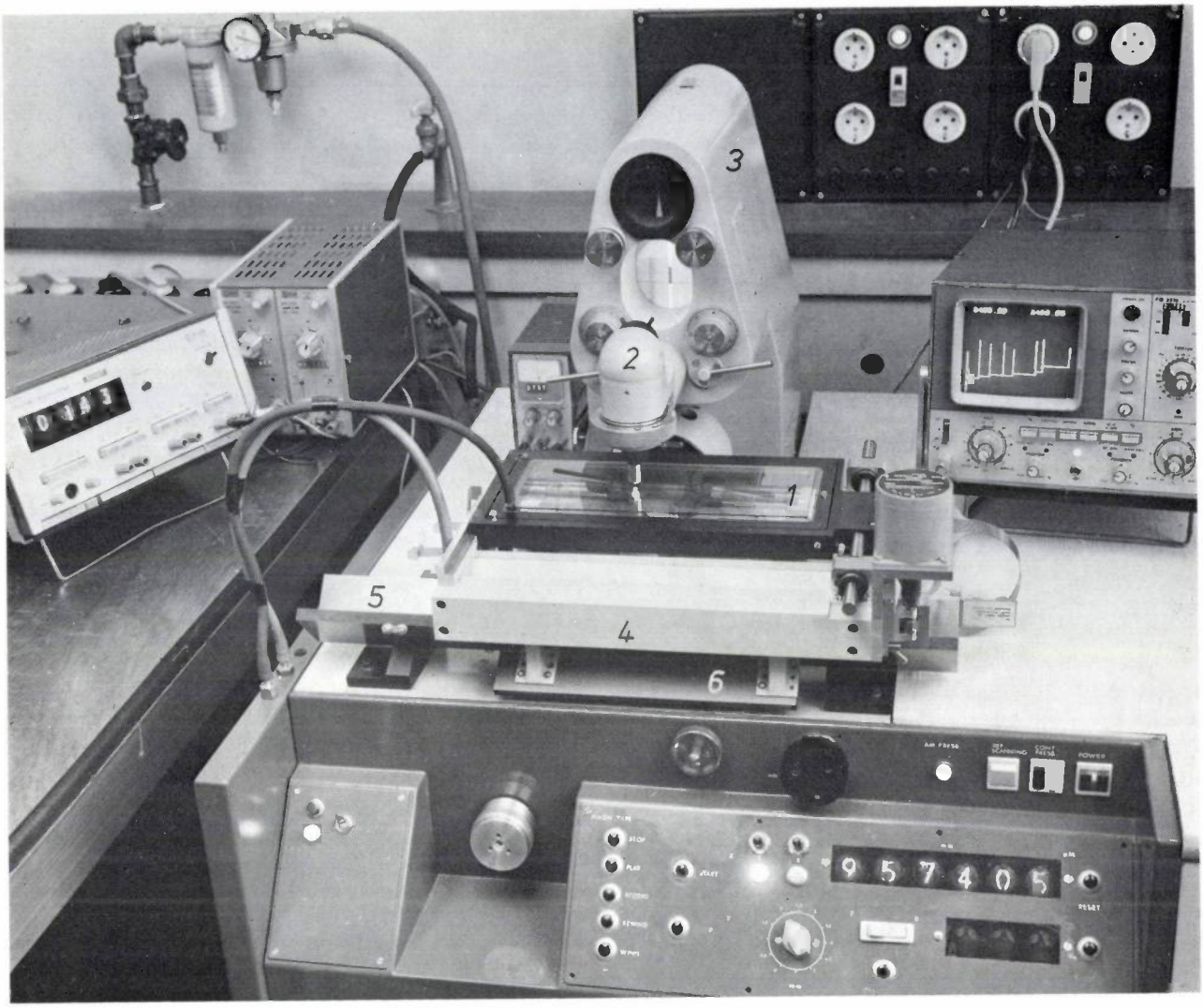


Fig. 1. Automatic microphotometer. A stationary beam of light passes upward through the photographic plate *1* (glass, 10×25 cm) and via the reading head *2* to a photosensitive cell in the housing *3*. A spectrum is scanned by moving the plate in the *x*-direction along the reading head by means of an electric motor. The slide *4*, on which the photographic plate is held by suction to a glass plate, travels on a cushion of air in a V-shaped guideway *5* (see fig. 2). The position is measured with an optical system, using a reflection phase grating mounted under the slide (*6*). To change from one spectrum to another the plate is moved in the *y*-direction. Both movements can be controlled by a computer as well as by hand. A digital reading of the position in the *x*-direction is given to an accuracy of 1.25 μm. Any part of the spectrum can be displayed on an oscilloscope screen. The measured blackening can be read at all times on the digital meter on the left. Parts of the 'Schnellphotometer' made by the Jenoptik company of Jena were used in the construction of this equipment.

magnetic bearings (see *fig. 2*) [8]. The displacement in the *x*-direction is measured by means of a built-in optical system designed and built at our laboratories [9]. It uses a reflection phase grating with a period of 10 μm, which measures positions to an accuracy of 1.25 μm.

The spectra are 25 cm in length; at each step of 5 μm a measurement is carried out and two items of information are fed to a computer: the exact position where the measurement is made and the value of the blackening. The measurement interval can be set to any value between 1 and 10 μm. It must be small enough to provide sufficient measurement points per spectrum line for the position of the peak to be determined [10]. On the other hand the interval must not be so small that there is not enough time for the first data reduction, which has to be executed by the computer during the scan, as we shall see below.

If the spectrum is a complicated one it may be desirable to check the results of a spectral recording. Once the spectrum has been stored in the computer memory, any part of it can therefore be retrieved and displayed on an oscilloscope screen.

Conversion of measured position into wavelength

The positions of the measurement points are expressed as a displacement in microns with respect to an arbitrary zero point, which is usually taken to be the centre of the strong carbon line at 2478 Å. In fact any line can be chosen as a reference, but the advantages of the C line are that it is readily recognizable, and that graphite electrodes are used for all the exposures. The micron scale for the positions must be converted into a wavelength scale, using the Ångström as the unit. The relation between these scales, called the *dispersion function*, depends on the spectrograph used for recording the spectrum; it can be derived both from the spectrum itself (internal calibration) and from an auxiliary spectrum (external calibration).

In internal calibration a number of reference lines of known wavelength and distributed over the whole spectrum are used. These lines are usually those associated with the main element in the sample. The positions of these lines are accurately measured and the dispersion function is derived from the results by the method of least squares. With the prism spectrographs that we used (Hilger, types E478 and E742) this function can be described by a fifth-degree polynomial. In internal calibration a series of reference lines is required for each type of sample and one element in each sample must therefore be known.

The first step in external calibration is to measure an iron spectrum; the dispersion function is next derived from this spectrum by means of internal calibration. This function is then used for a number of spectra recorded afterwards. Since the dispersion function can change during the exposures, repeated checks are necessary to make sure that it is still correct. An iron spectrum is therefore again recorded and the dispersion function again determined. With this procedure it is only necessary to have a series of iron reference lines available; nothing need be known about the composition of the samples.

In our automatic analysis procedure we use external calibration. The dispersion function is always checked after every nine spectra; special modifications have also been made that permit one dispersion function to be used for up to 54 spectra. The most important modification here was to the plateholder in which the photographic plate is inserted for recordings in the spectrograph. The plate should ideally take up a slightly curved, cylindrical position. In the original holder the plate was clamped in position by the lid along the two longer sides. In between, however, the plate could sometimes take up various unwanted positions in the form of a saddle surface. To avoid this a new lid was designed that gives exactly the right degree of curvature, with the plate held against it by suction from a vacuum (*fig. 3*). Other modifications include the improvement of the guideway along which the plateholder is moved between two exposures, and the use of temperature regulation in the room where the spectrograph is set up.

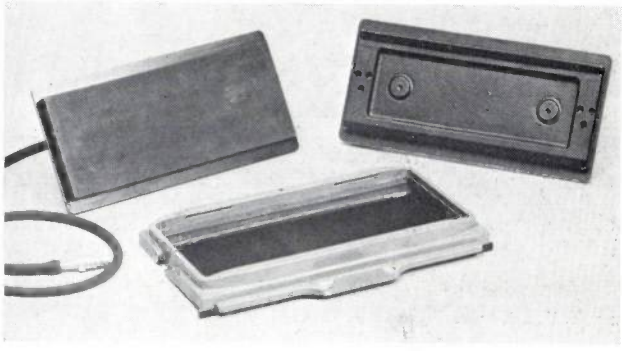


Fig. 3. Holder for the photographic plate. During the exposures in the spectrograph the plate must be given a cylindrical curvature in the holder. In the old model with the lid shown top right the plate was clamped top and bottom in such a way that it could sometimes assume the form of a saddle surface. In the new design (*left*) the lid has a number of narrow channels connected to a vacuum line, so that the whole of the plate is pulled against the lid and thus assumes the exact shape required.

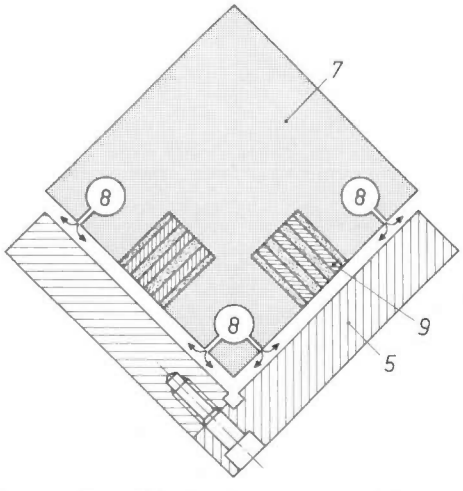


Fig. 2. Cross-section of the bearing system used for moving the plate slide in the x-direction. The slide is mounted on a guide block *7* that travels on a cushion of air in a V-shaped guideway *5*. The air is supplied through channels *8* at a pressure of *4* atmospheres. The distance between the block and the guideway is limited by two permanent magnets *9*, which give the system the required stiffness.

The determination of the dispersion function from 17 lines of an iron spectrum is carried out entirely by the computer. If any slight modifications in the dispersion polynomial are necessary at periodic recalibrations, these are also made by the computer without further need for human intervention.

[8] G. L. Walther, Magnetische rechtgeleiding, Polytechn. T. A **21**, 886A (also E **21**, 735E), 1966, and U.S. Patent No. 3.272.568, 13 Sept. 1966.

[9] H. de Lang, E. T. Ferguson and G. C. M. Schoenaker, Philips tech. Rev. **30**, 149, 1969.

[10] G. Brouwer and J. A. J. Jansen, Deconvolution method for identification of peaks in digitized spectra, Anal. Chem. **45**, 2239-2247, 1973 (No. 13).

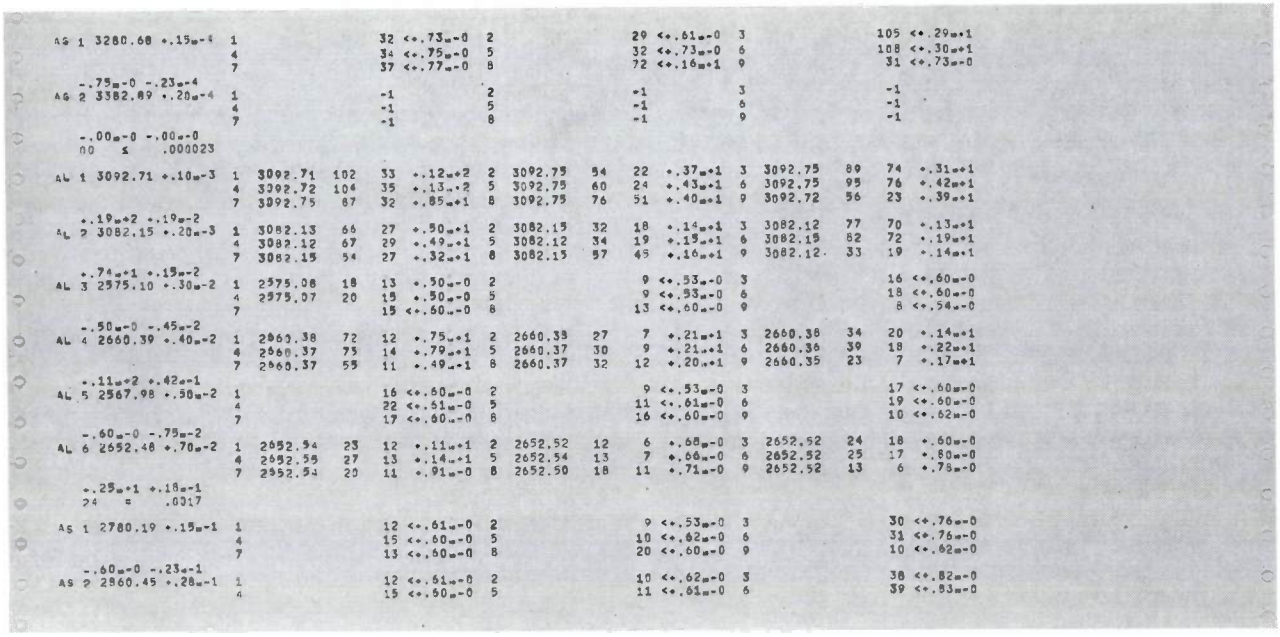


Fig. 4. Part of the list *DR*, giving the result of the second data reduction. The list relates to nine spectra recorded on one sample. It indicates for each of the 400 analysis lines whether the spectra contain a line that falls inside the associated wavelength window. If they do, the blackening values of line and background are given; if they do not, only the background blackening is given. After the number of the spectrum the following information appears: the measured wavelength of the line, the peak blackening, the background blackening and the net intensity, in that order. The figures *beside* the symbol for the element are the nominal wavelength and the sensitivity factor, and those *below* it are the average intensity and the provisional concentration, calculated entirely from data relating to that particular line.

Data reduction

With a measurement interval of 5 μm a spectrum 25 cm long provides data relating to 50 000 measuring points. It is impossible and unnecessary to store all this data in the computer memory. The amount of information gathered in the scanning of the spectrum is therefore reduced in such a way that only the blackening and the position of the peak of each detected line are retained, apart from the blackening of the associated background and the position where it was measured. The other measured values are not used. A line is regarded as having been detected when the peak blackening is above a certain threshold, which depends on the fluctuations in the background blackening immediately adjacent to the line.

After this first data reduction there remains on average data relating to 2000 lines per spectrum. This is still far more than is needed; the 400 analytical lines mentioned in connection with the conventional analysis, which are sufficient for determining 66 elements, are also sufficient for the automatic analysis. To enable further data reduction to be carried out the data for these 400 lines is stored in the computer memory. *Table I* shows this information for a few elements. A small tolerance region is reserved around each line wavelength, since the measured wavelengths do not always have exactly the correct value. The widths of these 'windows' have been empirically determined, and with the spectrographs mentioned above they are equal to 0.2 Å at a wavelength of 2500 Å and 0.4 Å at 3500 Å.

In the second data reduction the computer checks each of the 400 lines to determine whether the recorded spectrum available after the first data reduction contains a line inside the window. If it does, the blackening values of peak and background are noted, and if not, then the blackening of the background is noted. Lines

Table I. Some of the 400 analysis lines. The first column gives the elements, the second column gives the order for each line (a number indicating the intensity of the line: the strongest line of an element is of order 1), the third column gives the wavelength and the last column the sensitivity factor from which the concentration can be calculated.

Ag	'1'	3280.68	0.000015
	'2'	3382.89	0.00002
	'3'	2721.76	1
Al	'1'	3092.71	0.0001
	'2'	3082.15	0.0002
	'3'	2575.10	0.003
	'4'	2660.39	0.004
	'5'	2567.98	0.005
	'6'	2652.48	0.007
As	'1'	2780.19	0.015
	'2'	2860.45	0.028
	'3'	2744.99	0.05
	'4'	2898.71	0.14
Au	'1'	2675.95	0.0005
	'2'	3122.78	0.006
	'3'	2748.26	0.01

not inside any window are discarded. If a line does appear inside a window, its sensitivity factor is used to calculate a provisional concentration of the element to which it relates [11]. The blackening of the line first has to be converted into a light intensity, and as in the case of the conventional analysis described above, this is done with the aid of the emulsion calibration curve. The computer procedure used here is the Kaiser transformation [12]. For highly complex spectra, where some of the lines overlap, special methods are used to separate the peaks, but these will not be discussed here [13]. The computer notes the result of the second data reduction in a list *DR* (*fig. 4*). Since nine spectra are recorded for each sample the list contains the results of nine measurements for each of the 400 wavelength windows.

Qualitative and quantitative analysis

The actual analysis is based on the list *DR* remaining after the second data reduction, combined with the list of sensitivity factors (Table I) and with tables of interfering lines. The analysis is performed in a series of cycles of an iterative process; the procedure in the first two cycles is illustrated schematically in *fig. 5*.

In the first cycle all the lines of each element in the list *DR* are investigated one by one, starting with the strongest. The elements are sorted into 'definitely absent' and 'possibly present'. An element whose strongest lines are not present in the spectrum presents no difficulty at all. It is evident that the sample does not contain this element, whose lines cannot therefore interfere with other lines. In such cases it is important to know the detection limit, and this is calculated with the aid of the sensitivity factor from the background blackening in the window of the strongest line [14].

If blackening is found in the windows of the strongest lines it does not necessarily prove that that element is present, since it may be due to lines of another element. However, since the other elements are not yet known at this stage, no decision can be taken as to the presence of that particular element, and it is therefore provisionally categorized as 'possibly present'. A provisional concentration is now calculated as an average of the provisional concentrations of the detected lines listed in *DR*. At this stage possible interferences are not taken into account; the concentration cannot therefore be regarded as correct until it can be shown that there is no interference with the lines. Calculations of concentrations are preferably based on a number of lines, but lines that yield very different concentrations are rejected because of the strong presumption of interference. Lines whose density is equal to the saturation blackening of the emulsion are of course also of no use in the calculations. The results of this first cycle, the provi-

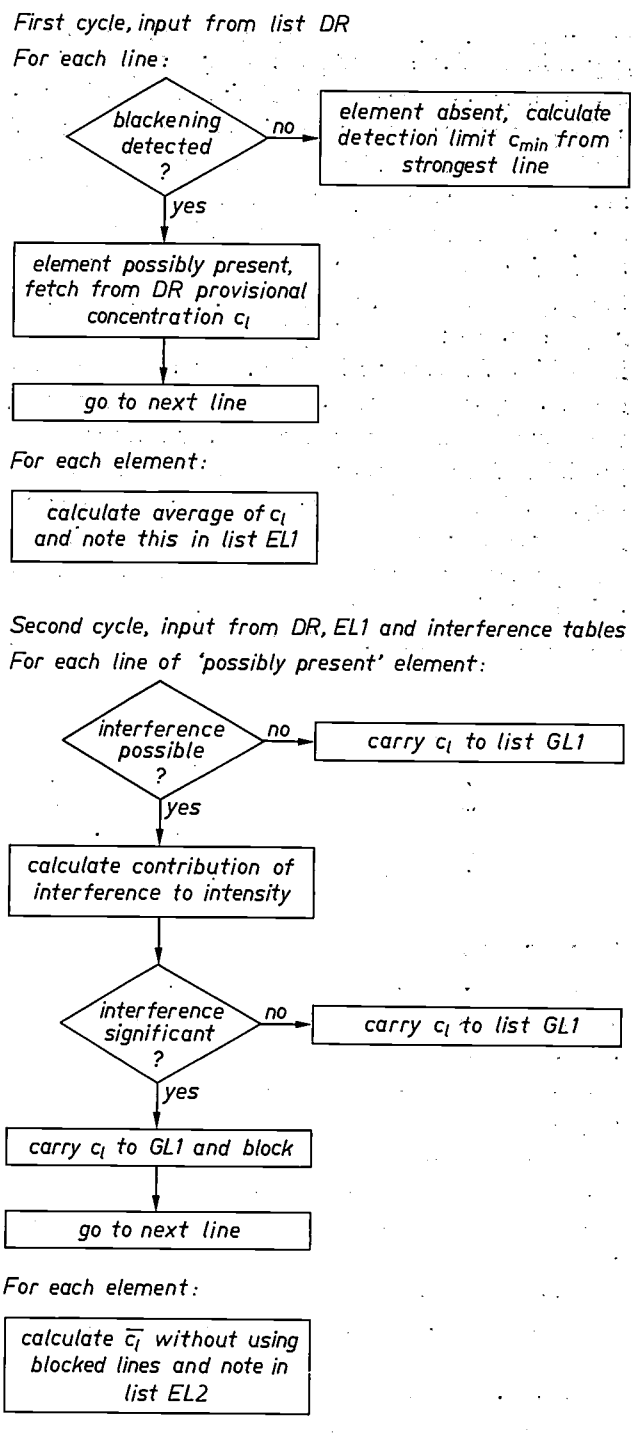


Fig. 5. Simplified flow diagram of the analysis. The second cycle is repeated until the list *EL* shows no further change.

[11] P. W. J. M. Boumans, Off-line data processing in emission spectrography, Colloquium Spectroscopium Internationale XVI, Part II, 247-253, Hilger, London 1971.

[12] H. Kaiser, Spectrochim. Acta **3**, 159, 1948.
M. Margoshes and S. D. Rasberry, Spectrochim. Acta **24 B**, 497, 1969.

[13] See the article of note [10].

[14] P. W. J. M. Boumans and F. J. M. J. Maessen, Evaluation of detection limits in emission spectroscopy, I, II, Z. anal. Chemie **220**, 241-260, 1966, and **225**, 98-107, 1967. See also the article of note [11].

sional concentrations of the various elements, are stored in an 'element list' *EL1*.

In the second cycle of the analysis the 'possibly present' elements are investigated further with the aid of the lists *DR* and *EL1*. The 'interference tables' stored in the computer memory are now used for eliminating the overlapping lines (see *Table II*). The interference tables give a complete list, for all the 400 windows of the analysis lines, of the lines of other elements that may be present in each window, and they also give the intensities of the interfering lines. For every line of an element that may be present a check is now made, from the known data, to determine whether interference is possible. If it appears from the tables that the only elements that could interfere with a line are those that have already been classified as 'definitely absent', then clearly interference cannot be a possibility and the calculated provisional concentration is at once transferred to a new list *GL1*, in which the result of the second cycle is noted for each line. If a line can be overlapped by lines of elements that are 'possibly present', then the provisional concentrations already calculated and the sensitivity factors of the interfering lines are used for calculating the contribution of the possible interference to the total intensity. If the contribution is found to exceed a specified threshold value, the line is 'blocked' (i.e. it is held in abeyance); if the interference is less than the threshold, it is then disregarded and the concentration is transferred to the list *GL1*. The threshold value is an empirically determined fraction of the total intensity, e.g. 10%.

For elements with blocked lines a new provisional concentration is now calculated without taking the blocked lines into account. In addition to the list *GL1*, the result of the second cycle thus includes a new element list *EL2*, giving the provisional concentrations of

— elements that are 'definitely absent', and cannot therefore cause interference,
— elements that are 'definitely present', and can therefore cause interference, and
— elements that are 'possibly present', and are thus possible sources of interference.

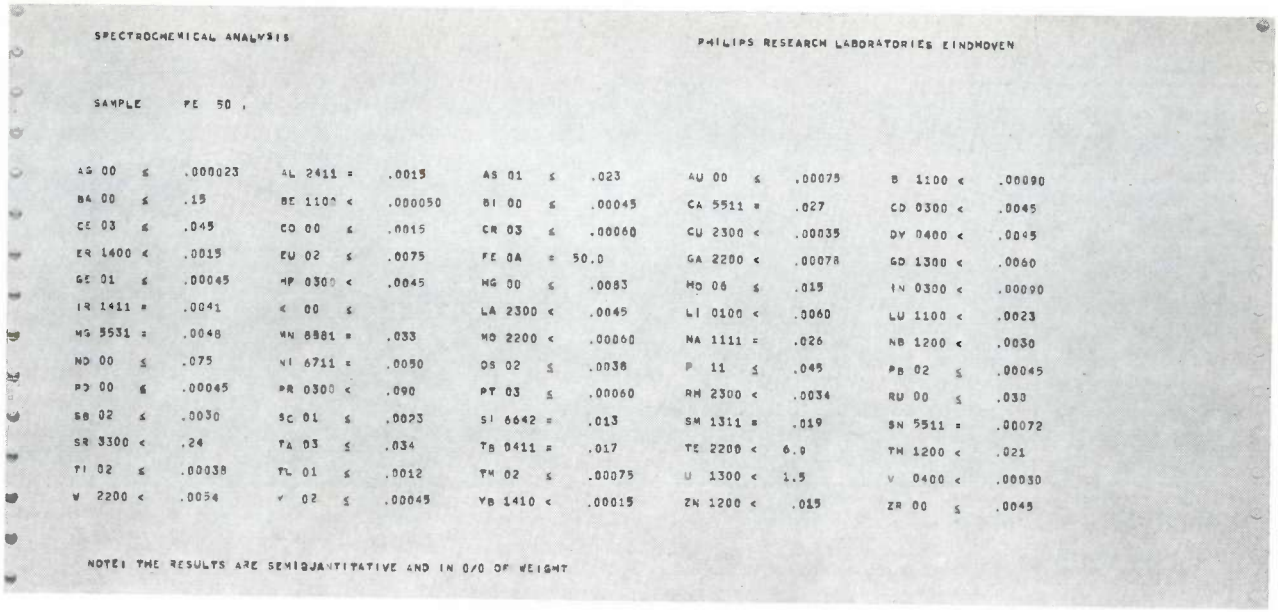
The procedure of the second cycle is now repeated, starting from the lists *GL1* and *EL2*. This time, however, a number of elements are known from the beginning to be 'definitely present' and it can therefore be established with greater certainty whether or not a line is subject to interference. This results in the lists *GL2* and *EL3*, in which more elements are now 'definitely present' and the concentrations have been calculated from more reliable data. The procedure of the second cycle is now repeated in an iterative process until it is established with certainty that all elements are either present or absent and until the same values for the concentrations are found in two successive lists *EL*. The final list *EL*, an example of which is shown in *fig. 6*, gives the ultimate result of the analysis. The number of cycles in the iterative process can be reduced if it is stated beforehand which elements are known to be definitely present and may therefore cause interference. This can often be done as for the main constituents, since a sample is seldom completely unknown.

Table II. Interference table for the strongest Ag line at 3280.68 Å. This lies in a wavelength window in which 22 elements can give interference lines. The first column gives the interfering element, the second the wavelength of the interfering line and the last column the value of $\sqrt{(10/q)}$ as a measure of the intensity of this line.

Th	3280.37	32
Ti	3280.39	2
U	3280.40	3
Ce	3280.48	32
Lu	3280.50	17
Rh	3280.55	10
U	3280.61	2
Ce	3280.67	10
Mo	3280.67	3
Pd	3280.68	3
Co	3280.68	3
Eu	3280.68	55
Cu	3280.68	2
In	3280.69	3
Th	3280.74	3
Zr	3280.75	3
Mn	3280.76	173
Sm	3280.84	32
Ta	3280.87	3
Mo	3280.88	3
Y	3280.91	10
Os	3280.91	3

It is obvious that large amounts of data have to be available for this procedure. Many analytical lines have to be known for each element, especially since some will have to be rejected because of interference. The tables of interfering lines must also be complete. Although most of this information can be found in standard works [15], the library of interfering lines and bands is not yet complete. What is more, the sensitivity factors of all these lines and bands have to be determined. It is therefore necessary to record and evaluate the spectra of a large number of elements. Even though the automatic microphotometer can be used for this purpose, much still remains to be done. In any case, the interference library will seldom be complete for the analysis of an entirely new type of sample. The analyst will therefore occasionally have to intervene in the iterative process. Here he can make use of the printouts of *DR* and the lists *GL* and *EL*, which are produced at the end of each cycle.

The calculations are performed by the Philips P9205 computer, which has a storage capacity of 16 000 words of 16 bits, extended with two PC1560 magnetic tape units, each with a capacity of 400 words per inch, and a P9242-030 disc store with a capacity of 120 000



SPECTROCHEMICAL ANALYSIS — PHILIPS RESEARCH LABORATORIES EINDHOVEN

SAMPLE FE 50 ,

AG 00	≤ .000023	AL 2411	= .0015	AS 01	≤ .023	AU 00	≤ .00075	B 1100	< .00090
BA 00	≤ .15	BE 1100	< .000050	BI 00	≤ .00045	CA 5511	= .027	CD 0300	< .0045
CE 03	≤ .045	CO 00	≤ .0015	CR 03	≤ .00060	CU 2300	< .00035	DY 0400	< .0045
ER 1400	< .0015	EU 02	≤ .0075	FE 0A	= 50.0	GA 2200	< .00078	GD 1300	< .0060
GE 01	≤ .00045	HF 0300	< .0045	HG 00	≤ .0083	HO 06	≤ .015	IN 0300	< .00090
IR 1411	= .0041	K 00	≤	LA 2300	< .0045	LI 0100	< .0060	LU 1100	< .0023
MG 5531	= .0048	MN 8881	= .033	MO 2200	< .00060	NA 1111	= .026	NB 1200	< .0030
ND 00	≤ .075	NI 6711	= .0050	OS 02	≤ .0038	P 11	≤ .045	PB 02	≤ .00045
PD 00	≤ .00045	PR 0300	< .090	PT 03	≤ .00060	RH 2300	< .0034	RU 00	≤ .030
SB 02	≤ .0030	SC 01	≤ .0023	SI 6642	= .013	SM 1311	= .019	SN 5511	= .00072
SR 3300	< .24	TA 03	≤ .034	TB 0411	= .017	TE 2200	< 6.0	TH 1200	< .021
TI 02	≤ .00038	TL 01	≤ .0012	TM 02	≤ .00075	U 1300	< 1.5	V 0400	< .00030
W 2200	< .0054	Y 02	≤ .00045	YB 1410	< .00015	ZN 1200	< .015	ZR 00	≤ .0045

NOTE: THE RESULTS ARE SEMIQUANTITATIVE AND IN 0/0 OF WEIGHT

Fig. 6. Final result of an analysis of an Fe sample. The quantities are expressed in percentages by weight. A code number after each element indicates the mode of detection. The element Mg, for example (seventh line, first element) is followed by 5531. The first 5 means 'the most sensitive five lines are present'; the second 5 means 'a line was found in five wavelength windows of Mg'; the 3 means 'out of the five lines found, three are free from interference'; the 1 means 'of the three free lines one was used for calculating the concentration'.

words. With this equipment a complete analysis of an unknown sample, including the read-out of the nine spectra, takes no longer than 45 minutes. The time taken by a conventional analysis depends very much on the nature of the sample. An uncomplicated sample can be analysed in about the same time as required by the automatic equipment, but the investigation of complex spectra can take several days. The gain in speed achieved through automation enables much more time to be devoted to the compilation of data and the study of matrix effects in new types of sample. These effects can be studied by recording the spectra of large numbers of different reference samples — a procedure that greatly improves the accuracy of the analysis.

Summary. An existing microphotometer, the 'Schnellphotometer' made by the Jenoptik company, has been modified so that emission spectra can be read fully automatically from photographic plates and the results fed to a Philips P9205 computer. The 25-cm long spectra are scanned at a speed of 1 cm/s. The blackening is measured at intervals of 5 μm, and the position at which each measurement is made is accurately determined (to < 1.25 μm) by an optical measuring system. The 50 000 items of data obtained in this way are reduced by the computer during the actual scanning; only the position and blackening of the peaks and the background blackening close to the lines are stored in the computer memory. The micron scale of the line positions is converted into a wavelength scale in ångströms; the dispersion function describing this relation is calculated from an iron spectrum (external calibration). To enable the dispersion function to be used for successively recorded spectra it was necessary to improve the positioning of the photographic plate in the spectrograph. The data is reduced a second time so as to limit the data to peak and background blackening in 400 wavelength windows. The lines in these windows can be used to detect 66 elements. The densities are converted into light intensities with the aid of the calibration curve of the photographic emulsion and from this input data, the computer calculates the concentrations of the elements present and the detection limits of those that are absent. Since lines of different elements may overlap, however, not all the data can be used for these calculations, and lines subject to such interference are rejected. This is done in a number of cycles in an iterative process that makes use of interference tables that list, for each wavelength window, the lines that can interfere. In this way a completely unknown sample can be analysed within 45 minutes.

[15] G. R. Harrison, Wavelength tables. M.I.T. Press, Cambridge, Mass., 1969.
J. A. Norris, Wavelength table of rare-earth elements and associated elements, Oak Ridge National Laboratory Publ. No. ORNL-2774.
R. W. B. Pearse and A. G. Gaydon, The identification of molecular spectra, 3rd edition, Chapman & Hall, London 1965.

Philips tech. Rev. 34, 330-338, 1974, No. 11/12

Neutron activation analysis

M.L. Verheijke

Introduction

In activation analysis a sample whose composition is to be determined is exposed to nuclear radiation, as a result of which radioactive nuclides are formed from the elements of the sample. From the spectrum of the radiation emitted during the decay of these nuclides a qualitative and quantitative determination can be made of the constituent elements of the sample. Although activation analysis is also carried out with charged particles such as protons, deuterons, alpha particles, ^{3}He particles and also with high-energy gamma quanta, most activation analyses are performed with thermal or 'moderated' neutrons (energy about 0.025 eV) [1]. For many elements the activation cross-section for thermal neutrons is so large that even very small quantities can clearly be detected. Epithermal neutrons (energies from about 0.5 eV to 1 MeV) and fast neutrons are less suitable for activation analysis because their activation cross-sections are much smaller. In any case, the available thermal flux during irradation in a nuclear reactor is many times greater than the epithermal flux.

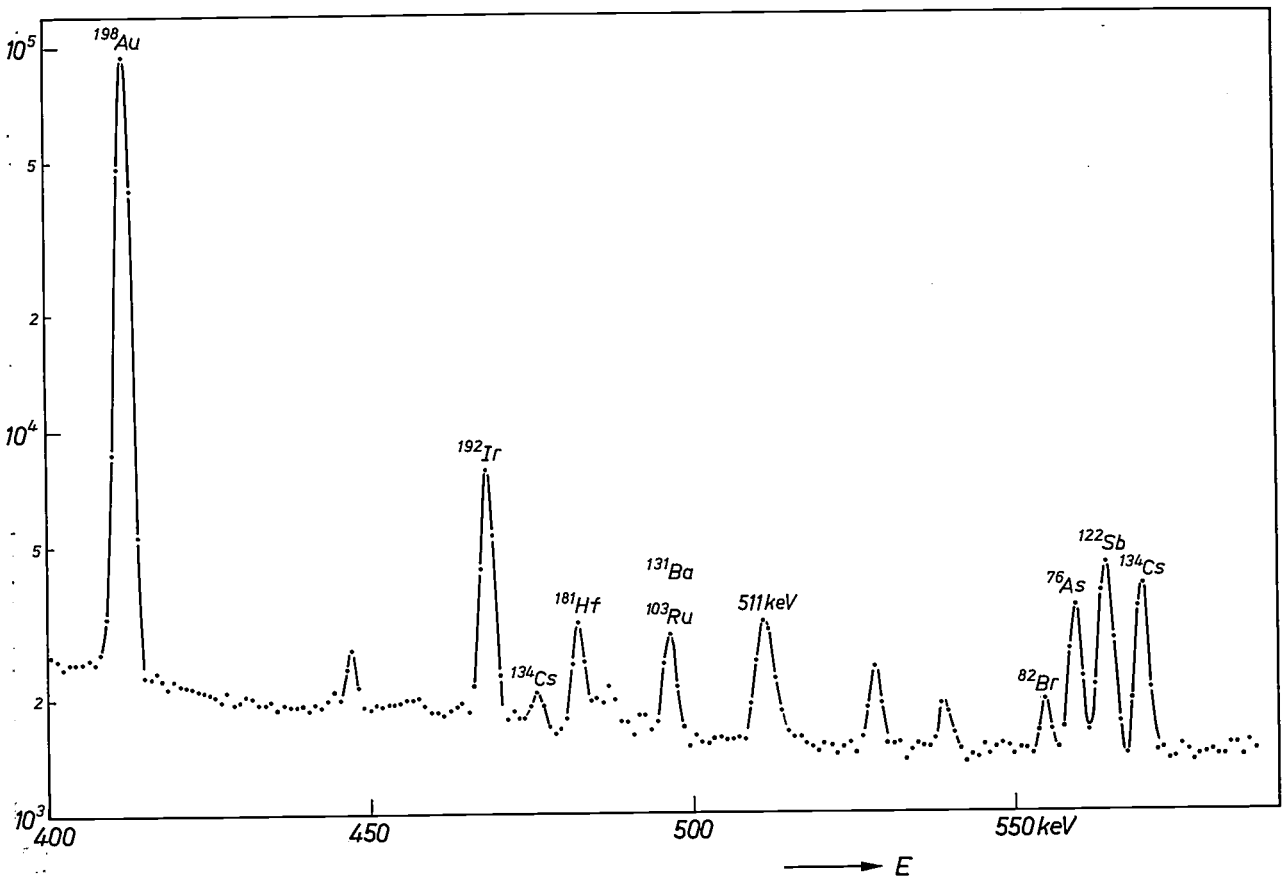


Fig. 1. Two small parts of a gamma spectrum, the whole of which covers the energy range between 70 and 2000 keV. The spectrum was recorded from a sample of very pure SiO_2 with trace impurities of the order of 0.01 to 100 ppb. Each dot gives the number of pulses in an energy interval of 1 keV. The peak at 511 keV is due to annihilation of β^+ and β^-particles.

When thermal neutrons are used, matrix effects can usually be neglected. They only become troublesome if the neutrons are so strongly absorbed by certain elements during the irradiation that the neutron intensity in the sample is no longer homogeneous. Elements that have a large activation cross-section for thermal neutrons — and which are thus in general

Ing. M. L. Verheijke is with Philips Research Laboratories, Eindhoven.

readily detectable by this method — are consequently not at all suitable as a matrix. This applies in particular to the elements Gd, Sm, Eu, Cd and Dy.

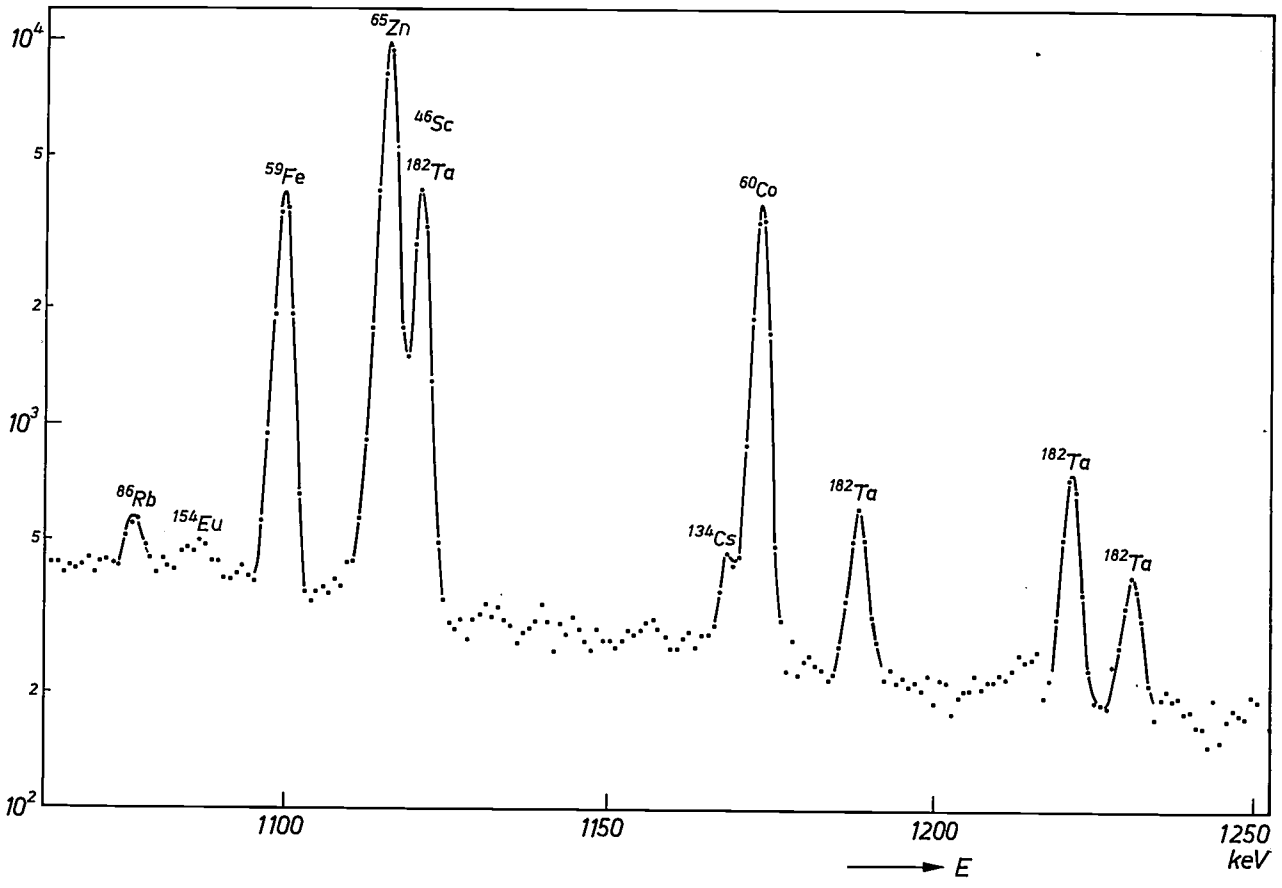
The main reaction that occurs in the irradiation of samples with thermal neutrons is what is called the (n,γ) reaction:

$$^{A}M_Z + n \rightarrow {}^{A+1}M_Z + \gamma\text{-quanta},$$

usually abbreviated to $^{A}M(n,\gamma)^{A+1}M$. The nuclide formed is then an isotope of the same element. Analysis is of course only possible if the nuclide is radioactive and emits gamma radiation during decay. (This gamma radiation should not be confused with the 'prompt' gamma radiation released during the (n,γ) reaction.) Since the spectrum of every gamma-radioactive isotope shows a number of peaks characteristic of each isotope, it is possible to deduce the constituent elements of an unknown sample directly from the peaks in the spectrum.

These spectra can be conveniently recorded by using a Ge(Li) detector [2] in combination with a pulse-height analyser (kick-sorter). The Ge(Li) detector has an energy resolution of 2-3 keV in a measuring range of 50-3000 keV. This high resolution permits gamma spectra to be recorded that enable the separate constituents to be distinguished in mixtures of a few tens of radionuclides (*fig. 1*). A condition is that none of the nuclides should have an activity many times greater than that of the others, for in that case the radiation of the other nuclides would be swamped in the Compton continuum of the strong emitter.

The heights of the peaks of the spectrum are not in themselves sufficient for quantitative analysis, since they depend not only on the concentrations of the constituent elements but also on such quantities as the activation cross-sections and the half-life of the nuclides formed from the reactions — quantities that vary considerably from one element to another. Quantitative analysis therefore requires a large number of calculations, and it is only the availability of the computer that enables the method to be used in multielement analysis.



[1] A standard work on neutron activation analysis is: D. de Soete, R. Gijbels and J. Hoste, Neutron activation analysis, Wiley-Interscience, London 1972.

[2] W. K. Hofker, Semiconductor detectors for ionizing radiation, Philips tech. Rev. 27, 323-336, 1966.

The method we have developed at Philips Research Laboratories also uses half-life differences for *qualitative* analysis. The spectrum of each sample is recorded at various intervals after the irradiation, the last recording being made as much as one month later. The whole procedure therefore takes rather a long time, but the uncertainty of the determination is much smaller than with only one recording.

For most elements neutron activation analysis is a highly sensitive method and is therefore eminently suited for trace analysis. By way of illustration *Table I* shows the detection limits for 58 elements obtained under ideal conditions. For some elements, such as Au, Eu and Sm, the limit is as low as 10^{-14} g, but some other elements, such as all the light elements up to Ne, can only be detected with difficulty if at all [3].

The accuracy of the calculated concentration is of course dependent on the nature of the element and the value of the concentration. It may also be affected by interference between peaks. The accuracy achieved in the most favourable cases is about 5%. In trace analysis this sometimes rather poor accuracy is not usually a disadvantage. However, if the method is to be used for determining higher concentrations than traces, and if at the same time a higher accuracy is desired, a standard must then be irradiated simultaneously with the sample and the spectra from both of them must be recorded [4].

In trace analysis conflict of course soon arises with the requirement that none of the nuclides should swamp the others during a recording of the gamma spectrum. During the measurement the matrix must therefore not be active, or only weakly so; in other words it must only contain weakly active or short-lived nuclides. This requirement is met by all the light elements up to Ne, and also by Mg, Si, Ti, Rh, V and Pb. A too-active matrix will have to be separated before the gamma spectrum is recorded by one of the usual chemical methods, such as extraction, ion exchange, distillation or precipitation [5].

Before describing the method developed at Philips Research Laboratories for carrying out qualitative and quantitative analyses, we should look briefly at the theoretical basis of quantitative analysis and the practical consequences derived from the equations involved.

Table I. Detection limits for 58 elements determined by neutron activation analysis under ideal conditions, i.e. no interference, high neutron flux (2×10^{14} cm^{-2}s^{-1}), irradiation time 24 hours, sample size 0.5 g, time since the end of irradiation 10 hours, measuring time 64 hours, large detector (108 cm^3). The table gives the contents that are only just detectable, in pp10^{12}.

Ag	8	Hf	1.5	Re	0.06
Ar	40	Hg	4	Ru	7
As	0.08	Ho	0.5	Sb	0.09
Au	0.004	In	20	Sc	0.15
Ba	150	Ir	0.02	Se	9
Br	0.15	K	10	Si	1.5×10^6
Ca	10 000	La	0.05	Sm	0.006
Cd	5	Lu	0.05	Sn	300
Ce	4	Mn	0.6	Sr	80
Co	3	Mo	2	Ta	3
Cr	10	Na	0.2	Tb	0.9
Cs	2	Nd	40	Te	9
Cu	2	Ni	700	Th	0.9
Dy	2	Os	1	Tm	1.5
Er	0.8	Pb	1×10^5	U	0.09
Eu	0.007	Pd	4	W	0.1
Fe	1000	Pr	1	Yb	0.15
Ga	0.2	Pt	2	Zn	10
Gd	2	Rb	40	Zr	80
Ge	40				

Theoretical basis of quantitative analysis

When an element consists of several stable isotopes, each of them may form a radioactive isotope by the capture of a neutron. Thus, from antimony with ^{121}Sb and ^{123}Sb the isotopes ^{122}Sb and ^{124}Sb can be formed. Although in principle only one isotope need be measured to determine an element (since the isotope percentages, the isotopic 'abundances', are constant) it is nevertheless the general practice to determine all the isotopes that have sufficient activity. It depends on the presence of interfering nuclides whether one particular isotope rather than another will lead to the most accurate result or the lowest detection limit.

The radioactivity of a particular isotope, that is to say the number of atoms D disintegrating per second at the end of the irradiation, is given by:

$$D = \frac{6.02\times10^{23} W\theta\, \Phi_{th}\, \sigma_{eff}}{A}(1 - e^{-\lambda t_1}). \qquad (1)$$

In this equation W is the quantity of the irradiated element (in g), θ is the abundance of the stable isotope considered, A the atomic weight of the element, Φ_{th} the thermal neutron flux, σ_{eff} the effective activation cross-section for the (n,γ) reaction giving rise to the relevant nuclide, λ the disintegration constant of the nuclide ($\lambda = \ln 2/T_{\frac{1}{2}}$, where $T_{\frac{1}{2}}$ is the half-life) and t_1 is the irradiation time.

Equation (1) contains an effective instead of an ordinary activation cross-section because it is not permissible to neglect the contribution of the epithermal neutrons. Although for most (n,γ) reactions the activation cross-section in the epithermal region is smaller than in the thermal region, it nevertheless gives sharp, high peaks for some elements as a function of the neutron energy. The influence of the epithermal neutron flux Φ_{epi} in the reactor can be taken into account by calculating the effective activation cross-section in eq. (2) as follows [6]:

$$\sigma_{eff} = \sigma_{th}\left(1 + \frac{\Phi_{epi}}{\Phi_{th}}\frac{I}{\sigma_{th}}\right). \qquad (2)$$

Here σ_{th} is the activation cross-section for thermal neutrons for the appropriate (n,γ) reaction and I is the resonance integral, which is a practical approximation to the activation cross-section for epithermal neutrons for this reaction.

The factor Φ_{epi}/Φ_{th} depends on the location in the reactor where the irradiation takes place. In practice this factor is not determined directly but is found from the cadmium ratio R for a suitable element (usually cobalt). This is done by irradiating two samples of this element, one unscreened and the other enclosed by 0.5 mm of cadmium. Since cadmium absorbs nearly all the thermal neutrons and allows the epithermal neutrons to pass through, the ratio R of the gamma activities of the samples after the irradiation is a measure of the ratio of the total flux to the thermal flux. This yields:

$$\frac{\Phi_{th}}{\Phi_{epi}} = (R-1)\frac{I}{\sigma_{th}}, \qquad (3)$$

where I and σ_{th} are the activation cross-sections of the element for which the cadmium ratio has been determined. Since the activation cross-sections of the various nuclides are known, the effective activation cross-section for any (n,γ) reaction at any location in the reactor can be calculated from the measured values of R.

Each peak in a gamma spectrum corresponds to gamma radiation of a certain energy. The area of the peak, representing the peak activity or *counting rate*, is a measure of the number of gamma photons of that energy, and hence of the activity of the radioactive nuclide emitting this radiation. The measured counting rate can be expressed in terms of the activity D of the nuclide immediately after irradiation as follows:

$$S = D\varepsilon a e^{-\lambda t_d}. \qquad (4)$$

Here S is the counting rate in the photopeak of energy E, t_d the time that has elapsed since the end of the irradiation, ε the peak efficiency of the detector for gamma photons of energy E, and a the abundance of this gamma photon during the decay of the nuclide ^{A+1}M.

When two or more gamma photons are emitted simultaneously during the decay of an atom the emission is referred to as a cascade. The simultaneous detection of these photons produces a pulse in the detector equal to the sum of the original pulses, resulting in a peak corresponding to the energy sum. The efficiencies for the original peaks are therefore smaller. A correction has been made for this in our computer programs, but will not be discussed here [7].

Some practical consequences

For a sensitive analysis it is of course necessary to have a high neutron flux Φ_{th} (see eq. 1) and a high peak efficiency $\varepsilon(E)$ (see eq. 4). Our experiments are usually carried out on the HFR nuclear reactor at Petten (the Netherlands) or the BR2 reactor at Mol (Belgium), both of which have a neutron flux of a few times 10^{14} cm^{-2}s^{-1}. For Ge(Li) detectors of about 100 cm^3, a type that we use, ε amounts to a few per cent at 1 MeV, and is roughly inversely proportional to E. As indicated in the introduction, the sensitivity differs from one element to another: σ_{eff} varies over several orders of magnitude. Even the abundance θ may not always be adequate in reactions that are otherwise perfectly usable: for example, the value of θ for the determination of Fe from ^{58}Fe(n,γ)^{59}Fe is only 0.0033. The same applies to the abundance of the gamma radiation: silicon is determined via ^{30}Si(n,γ)^{31}Si; the isotope ^{31}Si, however, is a virtually pure β^--emitter, and for the only gamma photon (1266 keV) the value of a is found to be only 7×10^{-4}.

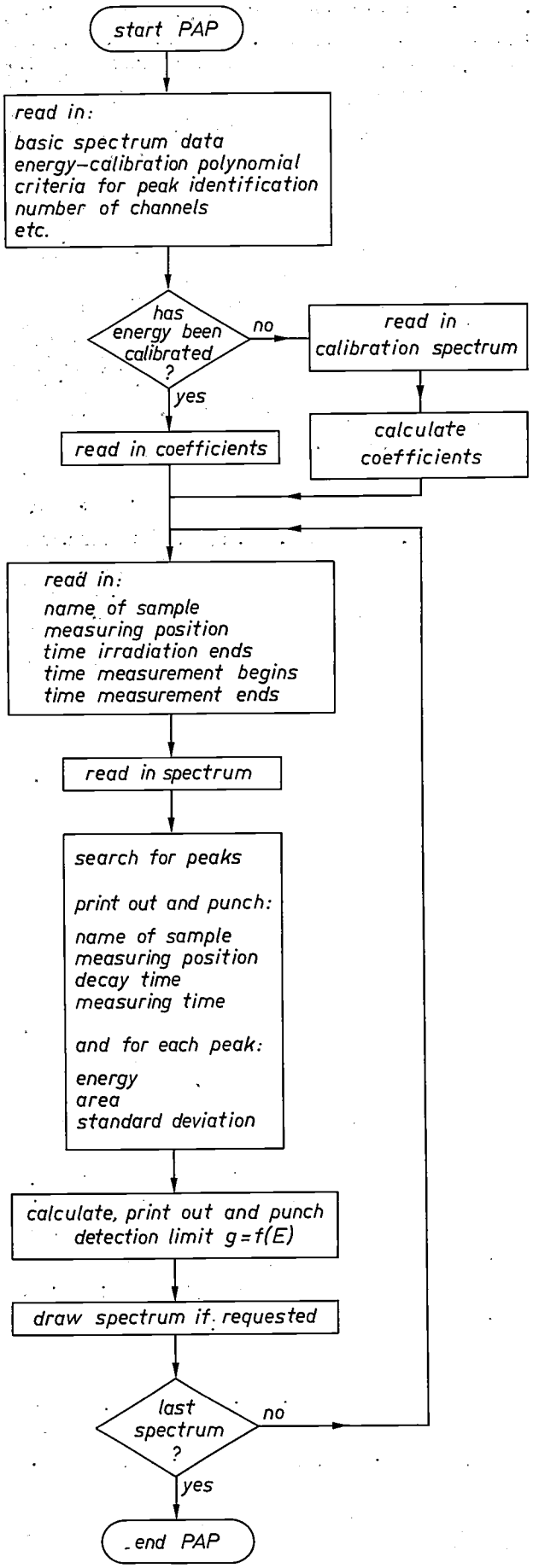
Another practical condition that must be imposed on a nuclide we wish to determine is that the half-life should not be too short ($e^{-\lambda t_d}$ small), because there is then no time for the measurement, nor should it be too long ($1 - e^{-\lambda t_i}$ small), because in that case the activity reached within the necessarily limited irradiation time is not high enough for the measurements to be carried out within a reasonable time.

The procedure in a complete analysis

The procedure adopted for analysing pure samples, in which there is little or no interference from the activity of the major constituents, is as follows. A few samples, for practical reasons no more than four, are weighed (about 0.1 to 0.5 g), wrapped in aluminium foil or enclosed in a silica tube, and then irradiated for 24 hours. The neutron flux present during the irradiation is determined with a flux monitor, consisting of a small quantity of Co in the form of Al wire with 0.1% Co, which is irradiated together with the samples.

After removing the wrapping, the spectrum of each sample is measured at four different times: 3 hours, 1 day, 1 week and 1 month after the irradiation. In certain circumstances (for example if chemical separa-

[3] Some elements that can only be detected from short-lived nuclides ($T_{1/2} < 1$ hour) are not given in the table, since it is based on spectra recorded ten hours after the irradiation.
[4] See the article by E. Bruninx and L. C. Bastings in this issue, page 350.
[5] J. C. Verplanke and P. N. Kuin, J. radioanal. Chem. **16**, 57, 1973.
P. C. M. N. Bruijs, J. radioanal. Chem. **16**, 115, 1973.
[6] O. T. Høgdahl, in: Radiochemical methods of analysis, Proc. Symp. IAEA, Salzburg 1964, vol. I, p. 23.
[7] J. C. Verplanke, Nucl. Instr. Meth. **96**, 557, 1971.



tion of the matrix is necessary) the first measurement is omitted because of lack of time. This means of course that the sensitivity of the analysis is then not so good for the short-lived nuclides. The four measurements on each sample enable us to identify the elements by using the half-lives as well as the energies of the peaks; this increases the reliability of the qualitative analysis. The spectrum emitted by the flux monitor is also recorded (in duplicate), and the computer then calculates the neutron flux from the intensities of the photopeaks of ^{60}Co at 1173 and 1333 keV from equations (1) and (4). In this calculation the amount of the element W is the known quantity in equation (1) and the flux Φ_{th} is the unknown.

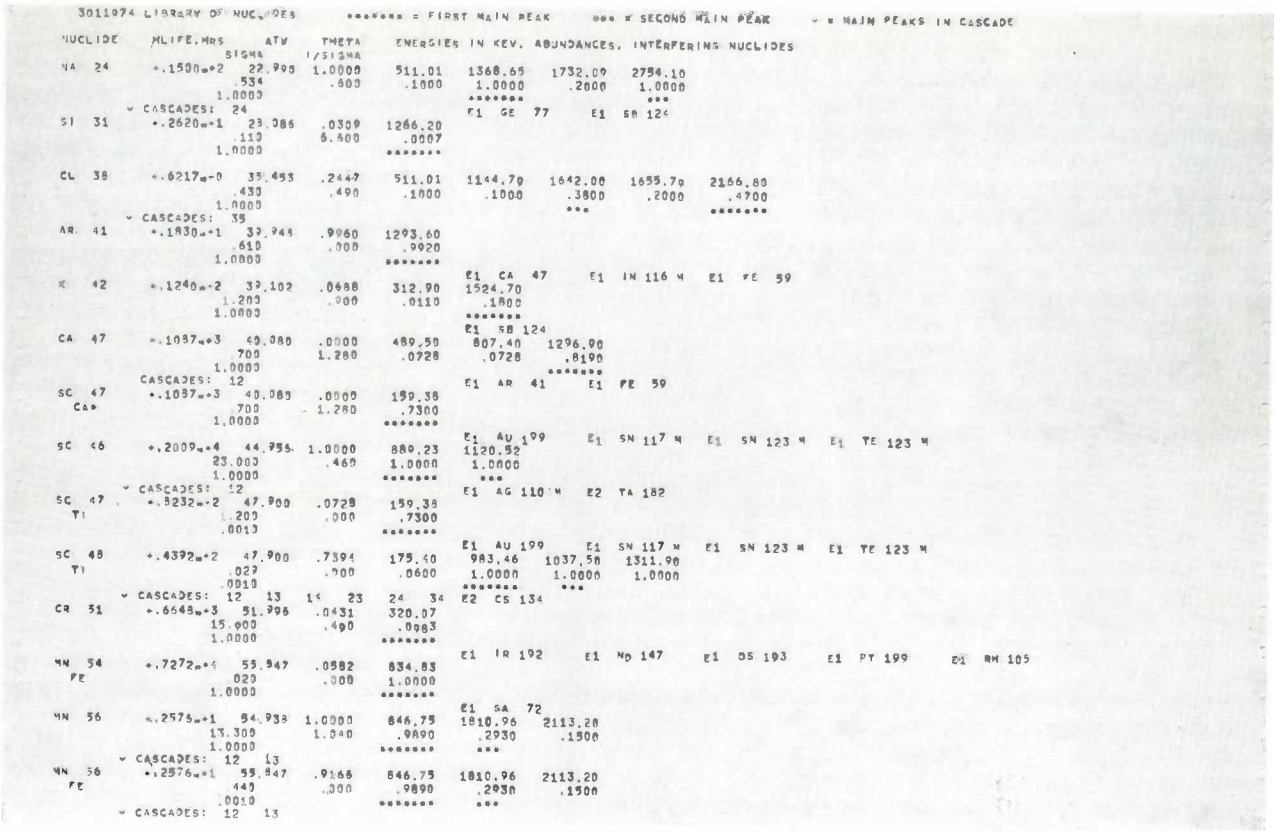
As mentioned above, the spectra are recorded with a Ge(Li) detector combined with a multichannel pulse-height analyser (kick-sorter), which sorts the pulses from the detector on the 2048 channels that cover the energy range from about 70 to 2000 keV. The energy scale of the whole system is calibrated with a spectrum recorded from a mixture of nuclides with accurately known gamma peaks. The calibration curve calculated by the computer from these known peaks is almost linear and is usually represented as a third-degree polynomial. This calibration only has to be made once a week.

Searching for the peaks

The first operation carried out with the computer is a search for all the significant peaks in the spectra with the aid of the computer program PAP (photopeak analysis program), see *fig. 2*. All programs are written in ALGOL 60 and run on the Philips Electrologica X8 computer. First a number of basic data of the various spectra are fed into the computer, such as the name of the sample, the times marking the end of the irradiation and the beginning and end of the measurement, and a code for the measuring conditions. We use three different detectors (Philips type APY with volumes of 43, 67 and 108 cm^3), with known peak efficiencies as a function of energy, and 13 standard 'measuring positions' per detector. These are the different distances to the detector and the various sample volumes ('point', 5 ml and 15 ml in standard vials). Data relating to the calibration of the energy scale, criteria for identifying the peaks, and so on are also required.

The program begins with a check on whether the energy scale has been calibrated, and if necessary the calibration is then carried out immediately. Next the

Fig. 2. Flow diagram of the computer program PAP (photopeak analysis program) for the searching for photopeaks. (The time between the end of the irradiation and the beginning of the measurement is called the 'decay time'.)



3011074 LIBRARY OF NUCLIDES ***** = FIRST MAIN PEAK *** = SECOND MAIN PEAK ~ = MAIN PEAKS IN CASCADE

NUCLIDE	HLIFE,HRS	ATW	THETA	ENERGIES IN KEV, ABUNDANCES, INTERFERING NUCLIDES			
	SIGMA	I/SIGMA					
NA 24	+.1500E+2	22.990	1.0000	511.01	1368.65	1732.09	2754.10
	.534		.600	.1000	1.0000	.2000	1.0000
	1.0000				*******		***
	~ CASCADES: 24				E1 GE 77	E1 SB 124	
SI 31	+.2620E+1	28.086	.0309	1266.20			
	.119		6.400	.0007			
	1.0000			*******			
CL 38	+.6217E-0	35.453	.2447	511.01	1144.79	1642.00	1655.79	2166.80
	.430		.490	.1000	.1000	.3800	.2000	.4700
	1.0000					***		*******
	~ CASCADES: 39							
AR 41	+.1830E+1	39.948	.9960	1293.60			
	.610		.000	.9920			
	1.0000			*******			
K 42	+.1240E+2	39.102	.0688	312.90	E1 CA 47	E1 IN 116 M	E1 FE 59	
	1.200		.900	.0110	1524.70			
	1.0000				.1800			
					*******			
					E1 SB 124			
CA 47	+.1037E+3	40.080	.0000	489.50	807.40	1296.90		
	.700		1.280	.0728	.0728	.8190		
	1.0000					*******		
	CASCADES: 12				E1 AR 41	E1 FE 59		
SC 47	+.1037E+3	40.080	.0000	159.38				
CA*	.700		1.280	.7300				
	1.0000			*******				
SC 46	+.2009E+4	44.956	1.0000	889.23	E1 AU 199	E1 SN 117 M	E1 SN 123 M	E1 TE 123 M
	23.000		.460	1.0000	1120.52			
	1.0000			*******	1.0000			
	~ CASCADES: 12				***			
SC 47	+.5232E+2	47.900	.0728	159.38	E1 AG 110 M	E2 TA 182		
TI	1.200		.000	.7300				
	.0013			*******				
SC 48	+.4392E+2	47.900	.7394	175.40	E1 AU 199	E1 SN 117 M	E1 SN 123 M	E1 TE 123 M
TI	.029		.900	.0600	983.46	1037.50	1311.90	
	.0010				1.0000	1.0000	1.0000	
	~ CASCADES: 12 13 14 23 24 34				*******	***		
CR 51	+.6648E+3	51.996	.0431	320.07	E2 CS 134			
	15.900		.490	.0983				
	1.0000			*******				
MN 54	+.7272E+4	53.947	.0582	834.83	E1 IR 192	E1 ND 147	E1 OS 193	E1 PT 199	E1 RH 105
FE	.029		.000	1.0000				
	1.0000			*******				
MN 56	+.2575E+1	54.938	1.0000	846.75	E1 GA 72			
	13.300		1.040	.9890	1810.96	2113.20		
	1.0000			*******	.2930	.1500		
	~ CASCADES: 12 13				***			
MN 56	+.2576E+1	55.847	.9166	846.75	1810.96	2113.20		
FE	.449		.000	.9890	.2930	.1500		
	.0010			*******	***			
	~ CASCADES: 12 13							

Fig. 3. Part of the 'library' containing data relating to the (n,γ) reactions of elements that could be of interest in an analysis. The list gives the name of the nuclide, half-life, atomic weight, isotopic abundance, activation cross-section, the ratio of resonance integral and activation cross-section, indications of 'cascades' of gamma energies, and the energies of at the most ten gamma peaks per nuclide (where present) with their abundances and with indications of the main peaks. The names of nuclides that may cause interference by emitting gamma photons with energies close to the main peaks are shown beside each nuclide.

peaks of the various spectra are detected. The way in which this is done will not be discussed here [8]; all that need be said here is that three parameters are determined for each peak in a spectrum; these are the energy (from the channel number), the counting rate S from equation (4) (the peak area) and the standard deviation s_S, calculated from the measuring statistics (Poisson distribution).

Since the original spectra with 2048 channels (with a maximum number of 2^{23} pulses per channel) are not preserved, steps must be taken at this stage to calculate the detection limit. It is important to know this if it should appear in the final stage of the calculations that a particular element is not present. The detection limit g_i — an area of a peak in the ith channel which, on statistical grounds, would only just be visible — can be calculated from the Compton background remaining when all peaks of the spectrum have been 'shaved off'. Only the result will be given here:

$$g_i = \frac{3}{t}\sqrt{2b_i y_i(1 + b_i)}, \qquad (5)$$

where b_i is the half-width of the peak (in channels), y_i is the channel content of the 'shaved' spectrum and t the measuring time.

The logarithm of g as a function of the energy can be approximated by a fifth-degree polynomial; its six coefficients are added to the other data of the spectrum to enable the detection limits of the elements to be calculated in the final stage of the analysis.

The spectrum at this stage can if required be displayed by means of an x-y plotter.

As the final result of the PAP program, the computer supplies a list giving the basic data mentioned above, the peak parameters and the six coefficients for calculating the detection limits for each spectrum. This list is printed out by a line printer and punched on tape for further processing by the computer.

Qualitative and quantitative analysis

In the second stage the computer identifies the nuclides and then calculates the quantities of the different elements [9]. It does this on the basis of the

[8] See: M. L. Verheijke, J. radioanal. Chem. **10**, 299, 1972.
[9] M. L. Verheijke and J. C. Verplanke, J. radioanal. Chem. **15**, 509, 1973.

spectral data as punched on tape by the PAP program. In addition the computer is supplied with data relating to the irradiation of the samples, such as neutron fluxes, Cd ratios and irradiation times, together with a large amount of nuclear data relating to the (n,γ) reactions that might be required for the analysis. In our case there are 135 reactions of 60 elements. *Fig. 3* shows part of the 'library' built up in this way.

The NAAP program (neutron activation analysis program), see *fig. 4*, first reads in the input data, and then, starting with the first spectrum, identifies the nuclides to which the peaks could correspond when only their energy is taken into account. A further examination is then made for each nuclide to find out whether its presence is indicated by the most commonly abundant peaks. Some nuclides have only one useful peak for this purpose, such as ^{51}Cr and ^{198}Au, but most nuclides have more and some have as many as 30 or more.

When the preliminary identification has been completed for all spectra the results are assembled for each sample (usually four spectra). The computer now uses equation (4) to calculate from the peak areas S of the two main peaks in these spectra the activity D at the end of the irradiation. If the preliminary identification was correct, the activities D calculated in this way for the nuclide in question should be identical within certain statistically defined limits. This is where the final decision is made on the presence or absence of a nuclide in the sample. If a nuclide is present, an activity is calculated for each of the main peaks as the weighted mean of the activities in the four spectra. The standard deviations from the peak areas found are a measure of the weighting factors. All decisions that the computer has had to take, for example as to whether or not a peak, and later a nuclide, is present, are based on certain statistical criteria, to some extent empirically established.

Now that the activity D is known, the computer can use equation (1) to calculate the quantity of the irradiated element, W. The detection limit for each nuclide can also be calculated with the aid of the fifth-degree polynomial and equations (1) and (4). The lowest limit found is the significant one. In the case of a short-lived nuclide this limit will be given by the spectrum that was first recorded, and by the last in the case of a long-lived nuclide.

The end result is printed out by the computer in the form of a list giving the calculated contents and detection limits together with an indication for every peak as to whether or not a nuclide is present that emits gamma radiation of nearly the same energy. The presence of any such interfering nuclide has the effect of raising the detection limit. Since the content cannot

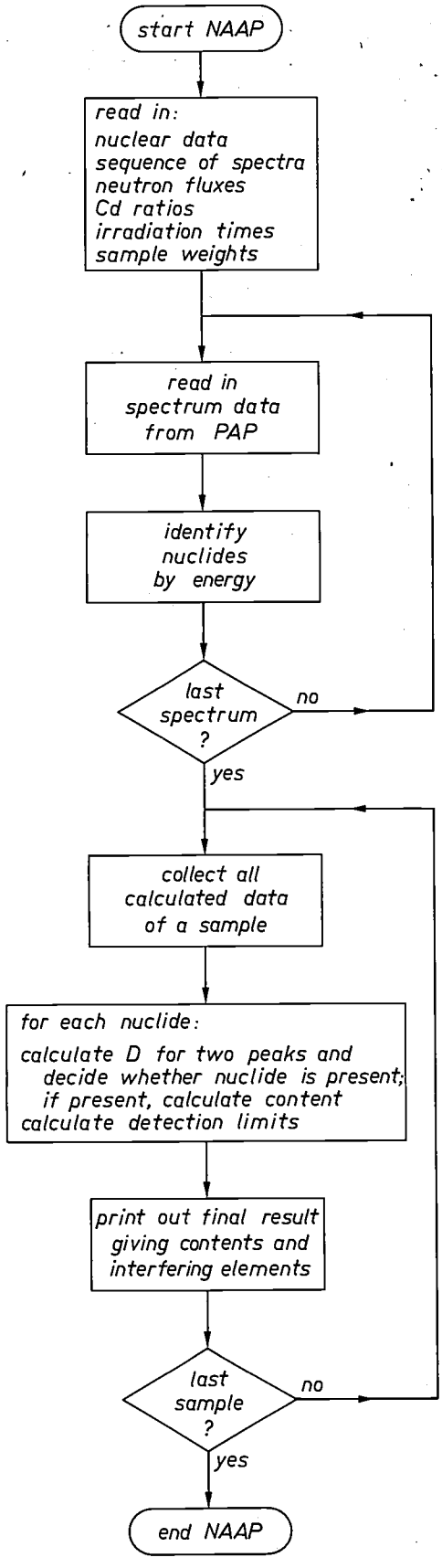
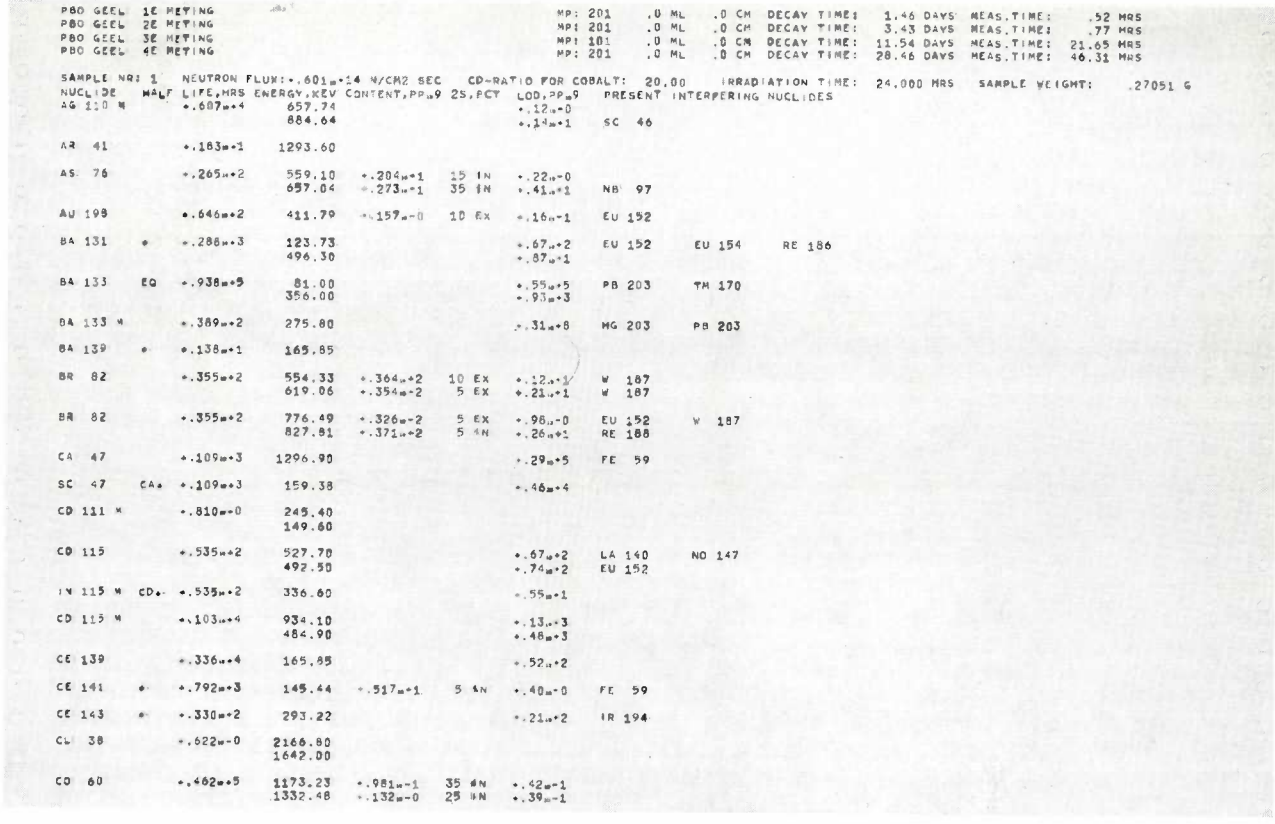


Fig. 4. Flow diagram of the computer program NAAP (neutron activation analysis program) for qualitative and quantitative analyses.



```
PBO GEEL  1E METING                                  MP: 201    .0 ML   .0 CM  DECAY TIME:   1.46 DAYS  MEAS.TIME:    .52 HRS
PBO GEEL  2E METING                                  MP: 201    .0 ML   .0 CM  DECAY TIME:   3.43 DAYS  MEAS.TIME:    .77 HRS
PBO GEEL  3E METING                                  MP: 101    .0 ML   .0 CM  DECAY TIME:  11.54 DAYS  MEAS.TIME:  21.65 HRS
PBO GEEL  4E METING                                  MP: 201    .0 ML   .0 CM  DECAY TIME:  28.46 DAYS  MEAS.TIME:  46.31 HRS

SAMPLE NR: 1   NEUTRON FLUX:+.601E+14 N/CM2 SEC   CD-RATIO FOR COBALT:  20.00    IRRADIATION TIME:  24.000 HRS   SAMPLE WEIGHT:    .27051 G
NUCLIDE   HALF LIFE,HRS ENERGY,KEV CONTENT,PPB 2S,PCT  LOD,PPB   PRESENT INTERFERING NUCLIDES
AG 110 M      +.687E+4    657.74                        +.12E-0
                          884.64                        +.14E+1   SC  46

AR  41        +.183E+1   1293.60

AS  76        +.265E+2    559.10    +.204E+1   15 IN    +.22E-0
                          657.04    +.273E+1   35 IN    +.41E+1   NB  97

AU 198        +.646E+2    411.79    +.157E-0   10 EX    +.16E-1   EU 152

BA 131   *    +.288E+3    123.73                        +.67E+2   EU 152    EU 154    RE 186
                          496.30                        +.87E+1

BA 133   EQ   +.938E+5     81.00                        +.55E+5   PB 203    TM 170
                          356.00                        +.93E+3

BA 133 M      +.389E+2    275.80                        +.31E+8   MG 203    PB 203

BA 139   *    +.138E+1    165.85

BR  82        +.355E+2    554.33    +.364E+2   10 EX    +.12E+1   W  187
                          619.06    +.354E+2    5 EX    +.21E+1   W  187

BR  82        +.355E+2    776.49    +.326E-2    5 EX    +.98E-0   EU 152    W  187
                          827.81    +.371E+2    5 IN    +.26E+1   RE 188

CA  47        +.109E+3   1296.90                        +.29E+5   FE  59

SC  47   CA+  +.109E+3    159.38                        +.46E+4

CD 111 M      +.810E-0    245.40
                          149.60

CD 115        +.535E+2    527.70                        +.67E+2   LA 140    NO 147
                          492.50                        +.74E+2   EU 152

IN 115 M  CD+ +.535E+2    336.60                        +.55E+1

CD 115 M      +.103E+4    934.10                        +.13E+3
                          484.90                        +.48E+3

CE 139        +.336E+4    165.85                        +.52E+2

CE 141   *    +.792E+3    145.44    +.517E+1    5 IN    +.40E-0   FE  59

CE 143   *    +.330E+2    293.22                        +.21E+2   IR 194

CL  38        +.622E-0   2166.80
                         1642.00

CO  60        +.462E+5   1173.23    +.981E-1   35 IN    +.42E-1
                         1332.48    +.132E-0   25 IN    +.39E-1
```

Fig. 5. Part of the final result of the analysis of a PbO sample. At the top some data are given for the four spectra recorded on this sample; the next line gives data relating to the sample. The results of the analysis for the various nuclides then follow (in alphabetical order), giving from left to right:
— the name of the nuclide (if the nuclide is not an isotope of the element in question the name of the element is printed after it),
— the half-life in hours,
— the energies, one below the other, of the two main peaks from which the contents were calculated,
— the contents, in this case expressed in ppb,
— the reproducibility of these contents ($2s$ in %) calculated from the spread in the four activities D for each peak and from Poisson statistics,
— the detection limit,
— the interfering nuclides if their presence is significant.

be corrected for interference, the analyst must later decide whether a content calculated from a peak subjected to such interference ought to be rejected. If two peaks of an element have been used, there still remains another peak with results that can be used. If there is a marked difference in the half-lives of two nuclides that interfere with one another, the computer takes the content for a short-lived nuclide from the spectrum that was first recorded, and for the long-lived nuclide from the last spectrum.

Fig. 5 shows the first part of the list with the final result of the analysis of a PbO sample. It can be seen, for example, that this sample contains arsenic with a content of 2.0 ppb (parts per billion, the usual abbreviation for 1 part in 10^9). The second peak of this element has clearly been affected by interference; this can be seen from the higher calculated content and the high detection limit. The quantity of bromine in the sample is determined from the nuclide ^{82}Br, and since the peaks in this case can be affected by quite a number of interfering nuclides, we have included this particular nuclide twice in the library, each with different main peaks. The good agreement between the four calculated contents (average 35 ppb $\pm$ 5%) enables us to conclude that, although interference is present, it has little effect in this sample. The results for the two peaks of cobalt differ fairly considerably, but this can be explained as being due to poor reproducibility, because the content is only a factor of 2.5 above the detection limit.

Owing to the fairly long time that elapsed before the first measurement could be made (1.46 days) — this was necessary to allow the complete decay of the nuclide ^{204m}Pb ($T_{1/2}$ = 67 min) — it was no longer

possible to detect the nuclides ^{41}Ar, ^{139}Ba, ^{111m}Cd and ^{38}Cl. The contents of altogether 18 elements were determined in this way and the detection limits of 35 elements. These detection limits are very important since the sample is of a substance required to have a high degree of purity; in such cases the guarantee that a sample does not contain more than a specific content of an element is just as valuable as an exact determination of that content.

Summary. After irradiation of a sample by thermal neutrons in a nuclear reactor the energy spectrum emitted by the sample is recorded with a Ge(Li) detector combined with a pulse-height analyser. With the locations and areas of the various peaks in this spectrum as input data, a computer carries out a qualitative analysis and calculates the contents of the constituent elements. The method is extremely sensitive (some elements can be detected in quantities as low as 10^{-14} g) and is therefore particularly suitable for trace analysis. The article gives the equations used in the calculations, describes the procedure of an analysis, and gives diagrams of the computer programs used for identifying a peak and calculating the quantities. A spectrum is recorded four times for each sample: 3 hours, 1 day, 1 week and 1 month after the irradiation. This makes it possible to use the half-lives in addition to the energies of the peaks for identifying the nuclides. The data relating to the (n,γ) reactions of different elements and the detectors used forms a 'library', which is stored in the computer memory. The neutron flux in the reactor is separately determined for each irradiation by irradiating a known quantity of cobalt together with the sample and measuring its activity. The cadmium ratio, which is a measure of the ratio of the flux of the thermal and epithermal neutrons, also has to be determined for all locations in the reactor. The result of the analysis is printed out on a list giving the contents of the detected elements in ppb, together with the detection limits of the elements that can be determined by this method.

Line-intensity calculations for X-ray fluorescence analysis

M. L. Verheijke and A. W. Witmer

X-ray fluorescence analysis is based on the emission of a characteristic X-ray spectrum when an element is irradiated with hard X-rays, gamma rays or fast electrons [1]. The relation between the concentration of the element and the intensity of the excited radiation is generally nonlinear, because of the 'matrix effect'. This is due to the presence of other elements in the sample, which absorb primary and fluorescent radiation and are often excited as well. In certain conditions the resultant radiation may also excite the element to be determined (secondary and tertiary excitation). Until a few years ago it was very difficult if not impossible to calculate this matrix effect, and an accurate X-ray fluorescence analysis required the use of a large number of reference samples. In this respect considerable progress has been made in recent years. The relation between concentration and intensity can now be fully calculated, although the number of elements cannot at present be more than three and the calculation has to be confined to the K lines. In this calculation the intensity is related to that of the pure element under the same experimental conditions.

The method that will be described here is known as the 'fundamental-parameter method' [2]. It is mainly used in laboratories where the nature of the samples sent for analysis varies considerably in a short time (other methods would require too many reference samples). In production-control laboratories, however, where long series of samples of the same type are analysed, it is more advantageous, and indeed more accurate, to use reference samples in the familiar way, or the 'empirical-coefficients method' [3], which is a kind of intermediate method.

The main purpose of our investigation was to check the reliability of the fundamental-parameter method experimentally [4]. We then used the method for calculating favourable experimental conditions for analyses using reference samples. Estimates of these conditions often turned out to be surprisingly far from the reality, demonstrating that calculations are by no means an unnecessary refinement.

For irradiating the samples we used both X-ray tubes and radionuclides. With the radionuclides we were able to use mono-energetic excitation, which simplified the calculations. The radiation excited in the specimens was analysed either with a Philips PW 1220 crystal spectrometer (using wavelength dispersion, see *fig. 1*), or with an Si(Li) detector, coupled to a 512-channel pulse-height analyser (using energy dispersion, see *fig. 2*).

Ing. M. L. Verheijke and A. W. Witmer are with Philips Research Laboratories, Eindhoven.

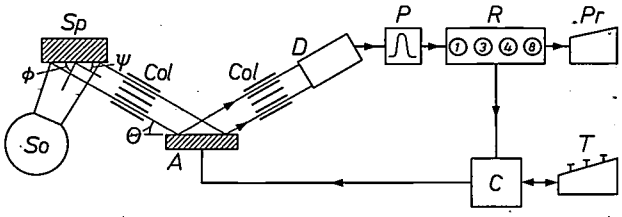


Fig. 1. Diagram of the X-ray fluorescence analysis equipment using wavelength dispersion. When the specimen *Sp* is excited with radiation from the source *So* (average angle of incidence ϕ) it emits fluorescent radiation, some of which passes through the collimator *Col* (take-off angle ψ) and strikes the analysing crystal monochromator *A* as a parallel beam. The reflected radiation is collected by the (proportional) detector *D*. Since only radiation whose wavelength satisfies Bragg's law is reflected, the angle of reflection Θ is a measure of the wavelength. A single-channel pulse-height discriminator *P* behind the detector *D* discriminates between radiation of different orders of reflection and eliminates background. *R* recorder. *Pr* printer. *C* computer for instrument control and evaluation of the measured data. *T* Teletype for data input and for recording the computed results.

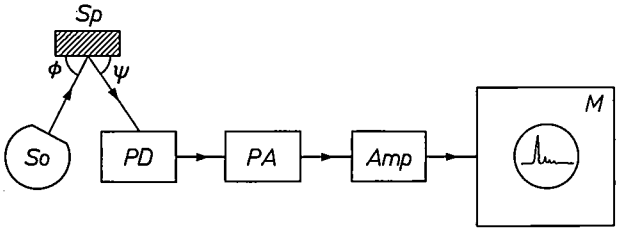


Fig. 2. Diagram of X-ray fluorescence analysis equipment using energy dispersion. *So* excitation source. *Sp* specimen. *PD* Si(Li) detector. *PA* preamplifier. *Amp* amplifier. *M* multichannel pulse-height analyser.

Calculated calibration curves for mono-energetic irradiation

We used the equations given in the Appendix to calculate the calibration curves (relative intensity as a function of concentration) for the AgKα, the SnKα and the SnKβ lines of the system Ag-Sn (*fig. 3*). The exciting radiation used was mono-energetic gamma radia-

[1] See for example: W. Parrish, Philips tech. Rev. **17**, 269, 1955/56; A. W. Witmer, Technische Rundschau **64**, No. 6, 25, 1972; R. Jenkins and J. L. de Vries, Practical X-ray spectrometry, Macmillan, London 1970.

[2] J. Sherman, Spectrochim. Acta **7**, 283, 1955 and **15**, 466, 1959. J. W. Criss and L. S. Birks, Anal. Chem. **40**, 1080, 1968. T. Shiraiwa and N. Fujino, X-ray Spectrom. **3**, 64, 1974.

[3] See W. K. de Jongh, X-ray Spectrom. **2**, 151, 1973. See also the article by Criss and Birks [2].

[4] The basic equations for the method, which have been taken from the literature, are given in the Appendix on page 342.

tion of 46.5 and 59.6 keV from the radionuclides ^{210}Pb (10 mCi) and ^{241}Am (1 mCi), respectively. Energy dispersion was used for the spectrum analysis.

As mentioned above, the advantage of monoenergetic excitation is that it simplifies the calculations. Since, for safety reasons, only fairly weak radioactive sources may be used, we had to rely on energy dispersion for the spectral analysis; in wavelength dispersion only a small part of the intensity of the radiation source is used.

In addition to the intensity of the SnKα line we also calculated the intensity of the SnKβ line, because the first line, which has an energy of 25.2 keV, is subject to interference from the AgKβ line (energy 25.1 keV), and this makes an experimental check difficult.

The reliability of the method was tested with the aid of standard samples containing 10, 30, 50, 60, 70 or 90 % of Ag and the corresponding quantities of Sn. The specimens were discs with a diameter of 30 mm and a thickness of 5 mm, which also met the usual requirements for surface smoothness and sample homogeneity in X-ray fluorescence analysis.

The relative intensity values measured are indicated in the graphs by circles and squares. It can be seen that there is satisfactory agreement between calculations and experimental results. Silver absorbs the SnKβ radiation much more strongly than the SnKα radiation, which explains the difference in the shape of the calibration curves for these two X-ray lines. It is also noteworthy that the calibration curves for the Kα lines of Ag and Sn are almost independent of the excitation energy, which is not the case for the SnKβ line.

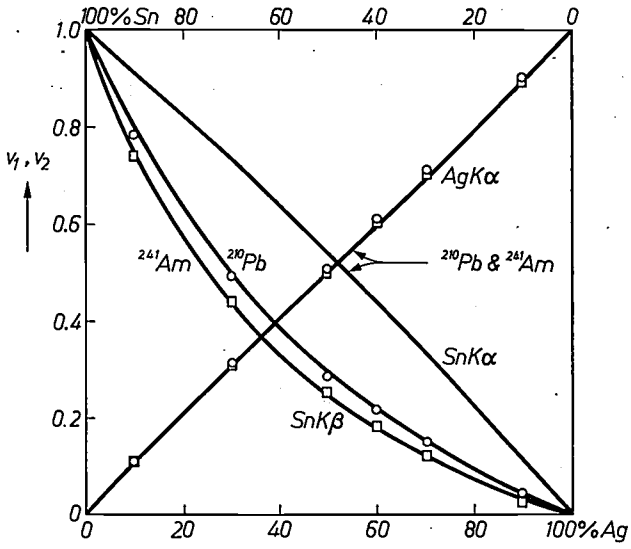


Fig. 3. Calculated calibration curves (relative intensity as a function of concentration in percentage by weight of the system Ag-Sn), experimentally verified. The exciting radiation came from ^{210}Pb and ^{241}Am (circles and squares respectively), corresponding to excitation energies of 46.5 and 59.6 keV. Energy dispersion was used for the spectral analysis. The angle of incidence ϕ was 64°, the take-off angle ψ was 84°.

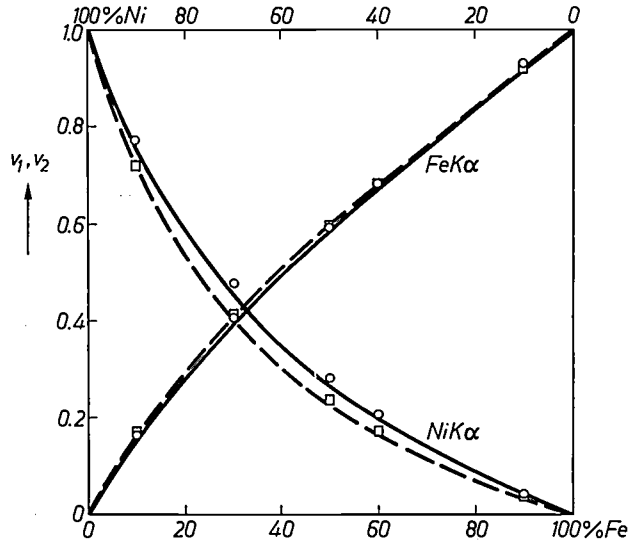


Fig. 4. Calculated and experimentally verified calibration curves for the system Fe-Ni with poly-energetic excitation and spectrum analysis by wavelength dispersion. The solid lines were calculated for excitation with a tungsten tube (50 kV); the circles represent the corresponding calibration points. The dashed curves were calculated for a chromium tube (45 kV); the squares represent the corresponding calibration points. $\phi = 65°, \psi = 35°$.

Calculated calibration curves for poly-energetic irradiation

Fig. 4 shows calculated calibration curves and experimentally determined calibration points for the system Fe-Ni. For these experiments poly-energetic excitation was used, obtained from a tungsten or chromium tube (50 or 45 kV respectively) combined with wavelength-dispersion equipment.

The dimensions of the standard specimens for the system Fe-Ni were the same as those for the system Ag-Sn. The specimens consisted of alloys containing 10, 30, 50, 60 or 90 % of iron. In spite of the more complicated calculations, here too the agreement between calculations and experimental results is satisfactory.

The NiKα line is strongly absorbed by iron, which explains the concave shape of the Ni curve. The slightly convex shape of the Fe curve is due to secondary excitation of iron by NiK radiation.

Finally, an investigation was made of a ternary alloy of Fe, Ni and Co, the familiar Fernico. The result is shown in *Table I*. The specimen was irradiated in four different ways and the intensities were measured with the aid of wavelength dispersive equipment. In comparing the results it should be borne in mind that there is secondary and tertiary excitation as well as primary excitation. Thus, iron has four different modes of excitation: direct excitation by the X-ray tube, secondary excitation by Co and Ni radiation, and tertiary

excitation by Co radiation which itself is excited by Ni radiation. Cobalt has two modes of excitation: primary by the tube and secondary by Ni radiation. Nickel is only subject to primary excitation by the tube.

All these effects had to be included in the calculation. Taking this into account, the agreement between the calculated and measured relative intensities can be considered satisfactory.

Tertiary excitation

It is often stated in the literature that the intensity due to tertiary excitation (i.e. primary energy excites element A which excites element B which, in turn, excites element C) constitutes only a small fraction of the total intensity of the radiation emitted by an element. To investigate this we have calculated for four different Cr-Fe-Ni alloys the ratio F of the tertiary intensity of the CrKα line with respect to the primary excited intensity, as a function of the excitation energy. The results are summarized in *fig. 6*.

The graph shows that cases can easily be found where the tertiary intensity is more than 10% of the primary intensity. If chromium is present in traces this value may even be as high as 20%, as can be seen from the upper curve. If the chromium concentration is low, the concentrations of the energy-transferring elements Fe and Ni are high. In this case the tertiary effect is evidently relatively strong.

We have also calculated (though the figures are not shown here) that in the case of a ternary alloy containing equal quantities of antimony and silver and less than 10% molybdenum, the tertiary intensity of the MoKα line may even exceed 40% of the primary intensity.

Table I. Calculated and experimentally determined relative intensities of the FeKα, CoKα and NiKα lines excited from Fernico (53.5% Fe, 17.7% Co, 28.8% Ni).

X-ray tube	FeKα		CoKα		NiKα	
	Calculated	Experimental	Calculated	Experimental	Calculated	Experimental
W-50 kV	0.584	0.578	0.186	0.172	0.150	0.150
Cr-45 kV	0.593	0.585	0.188	0.174	0.128	0.130
Mo-45 kV	0.590	0.580	0.187	0.173	0.122	0.118
W-20 kV	0.583	0.581	0.186	0.176	0.153	0.157

Relative intensity as a function of the excitation energy

Even when 'classical' calibration methods are used, with sufficient reference samples available at the right time, it may still be useful to calculate the relative intensities before starting the analysis. The processes of absorption and excitation in X-ray fluorescence are so complicated that this is the only way of obtaining an overall view of the effects and interactions of the various physical and instrumental parameters, thus making it possible to choose the best experimental conditions.

Fig. 5 shows the result of such a calculation. It gives a plot of the relative intensities of the CuKα and AgKα lines of an alloy containing 50% Cu and 50% Ag, as a function of the excitation energy.

It can be seen that in the first instance the relative intensity decreases with increasing excitation energy. This shape of curve would not at first sight have been expected, since the absolute intensity *increases* with increasing excitation energy. The explanation is to be found in the fact that the absorption by the pure element decreases more rapidly with increasing excitation energy than the absorption by the specimen (see equation (4) in the Appendix, page 343).

The calculation also shows that at energies higher than the AgK absorption edge, where the Ag radiation begins to become significant, secondary excitation of Cu by Ag radiation takes place.

A calculation of the type described above is useful because it enables the excitation energy to be determined at which the calibration curve has a suitable shape. It is desirable, for accuracy of the analysis, to have a calibration curve that shows the minimum deviation from linearity. For this reason the excitation should be carried out with an energy that corresponds — given the content of 50% Cu and 50% Ag — as far as possible to a relative intensity of 0.5.

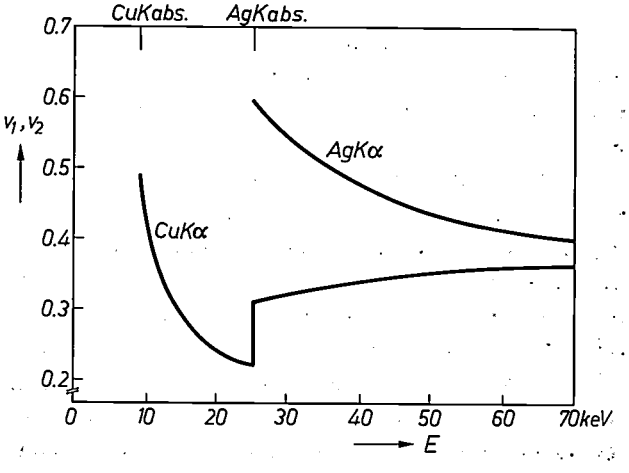


Fig. 5. Relative intensities of CuKα and AgKα lines as a function of the excitation energy E, calculated for an alloy consisting of 50% copper and 50% silver. The absorption edges are indicated above. $\phi = 65°$, $\psi = 35°$.

Effect of dilution on the relative intensity

Fig. 7 shows the results of calculations for the system Br-Cs in an aqueous solution, demonstrating how dilution with water affects the relative intensity of the X-ray lines. The solid curves $v_{Br}(0)$ and $v_{Br}(10)$ are calibration curves for BrKα lines; one applies to bromine in pure water (0% Cs) and the other to the same solutions but containing 10% Cs. It is evident that water (whose absorption incidentally is not negligible) does not absorb nearly as much as Cs, so that the intensities are higher in the system consisting only of Br and water. This difference in absorption is partly compensated by the secondary excitation of Br by Cs radiation. The dashed curve is the result of calculations of the intensity of the BrKα line, in which this excitation effect was not considered, so that the full absorption effect is shown.

The upper curve is particularly interesting. It gives the calculated relative intensity of the CsKα line for 10% Cs in an aqueous solution as a function of the bromine concentration. As can be seen, a 10% solution of an element can yield an intensity not much lower than that measured for the pure element. The consequences of dilution with water are to a large extent compensated by the low absorption of water. Thus, calculation demonstrates that solution in water is an aid that can be used more often than might have appeared likely without prior calculation.

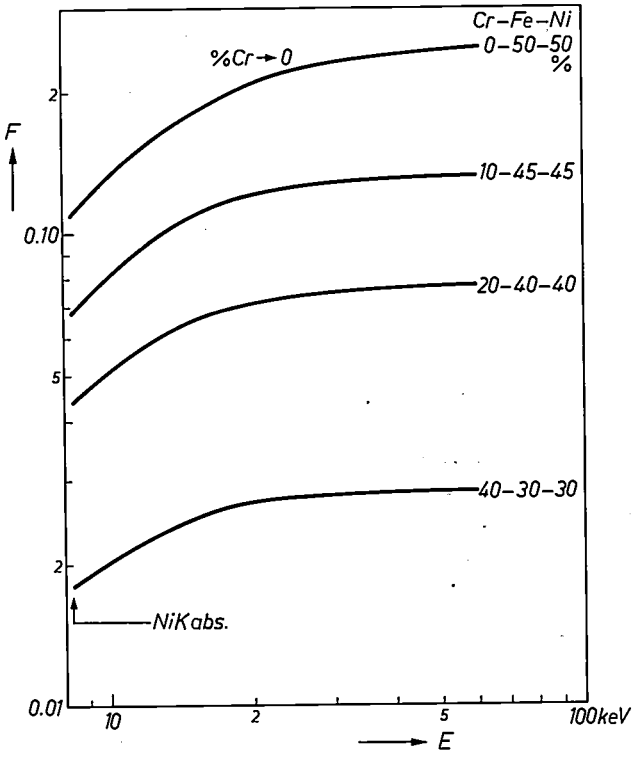


Fig. 6. The ratio F of the relative tertiary intensity to the relative primary intensity for the Kα line of Cr, calculated for four alloys of chromium, iron and nickel, as a function of the excitation energy E. The upper curve is calculated for a chromium percentage approximating to zero. $\phi = \psi = 45°$.

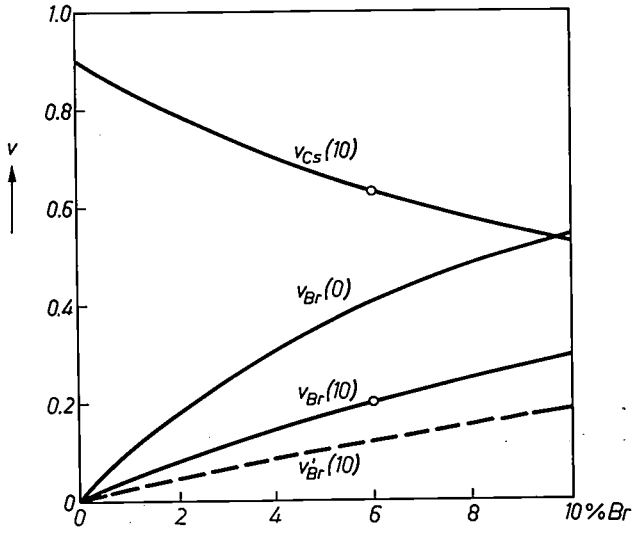


Fig. 7. Calculated relative intensity of the BrKα and CsKα lines of the system Cs-Br-H_2O as a function of the percentage by weight of Br, the first in the presence of 0% and 10% Cs, the second in the presence of 10% Cs. The secondary excitation of Br by Cs radiation was not taken into account in calculating the dashed curve. The circles relate to CsBr. $\phi = \psi = 45°$.

The figure shows that the substitution of water for Cs (0% Cs instead of 10% Cs) results in a marked increase in the relative intensity of the BrKα line, indicating that water absorbs much less than Cs. After correction for secondary excitation (dashed curve) an even greater difference in relative intensity is found. The upper curve shows that when the Cs content is reduced to 10% by dilution with water, the relative intensity of the CsKα line only decreases to 90% of that of the pure element.

Appendix. Method of calculation

We shall briefly describe here the calculation by the fundamental-parameter method of the relative intensities of X-ray lines generated in a ternary system. We shall first give the calculation for the case of mono-energetic excitation, and then for poly-energetic excitation.

The relative intensity of an X-ray line of an element is defined as the absolute intensity of the line divided by the absolute intensity of the same line of the same element in a pure state, measured under identical instrumental conditions.

The use of relative intensities eliminates the effect of various instrumental parameters, and also simplifies the calculations.

In the following we shall confine ourselves to the K lines of the elements. Since, when this restriction is made, an element can only be excited by a line of an element with a higher atomic number, we index the elements in order of increasing atomic number Z. Thus, in the case of a ternary system, the subscripts 1, 2 and 3 indicate the three components in the order $Z_1 < Z_2 < Z_3$. The subscript 4 will be used for the mono-energetic exciting radiation. Let v_i be the relative intensity of the radiation to be measured and w_i the concentration (weight fraction) of the element i. We then obtain:

$$v_3 = b_3 w_3, \tag{1}$$

$$v_2 = b_2 w_2 (1 + c_{432} w_3), \tag{2}$$

$$v_1 = b_1 w_1 (1 + c_{421} w_2 + c_{431} w_3 + c_{4321} w_3 w_2). \tag{3}$$

Here the factors b_i relate to the absorption and the factor c to the secondary and tertiary excitation.

The factors b_i depend on the degree of absorption of the exciting radiation and the radiation being measured. They are the solutions of the integrals with exponential absorption terms for an infinitely thick specimen:

$$b_i = \frac{\mu_{4i} + \mu_{ii}A}{\mu_{4\cdot} + \mu_{i\cdot}A} \quad (i = 1, 2, 3). \tag{4}$$

The first subscript of the mass-absorption coefficients μ relates to the absorbed excitation energy (in a given case the K energy of the indexed element) and the second relates to the absorbing medium. If we use a point instead of the second subscript we refer to the whole specimen, thus:

$$\mu_{i\cdot} = \sum_{j=1}^{3} w_j \mu_{ij}. \tag{5}$$

The letter A stands for $\mathrm{cosec}\psi/\mathrm{cosec}\phi$, where ϕ and ψ are the angles between the specimen surface and the incident and emergent beams respectively (figs. 1 and 2).

In equations (2) and (3) the correction terms in parentheses stand for secondary and tertiary excitation. The parameter c_{4jk} relates to the excitation of element k by energy 4 via element j ($j > k$) [2]:

$$c_{4jk} = \frac{\tau_{4j}K_j\omega_j t_{jk}}{2\tau_{4k}}\left\{\frac{1}{\mu'_{4\cdot}}\ln\left(1+\frac{\mu'_{4\cdot}}{\bar{\mu}_{j\cdot}}\right)+\frac{1}{\mu''_{k\cdot}}\ln\left(1+\frac{\mu''_{k\cdot}}{\bar{\mu}_{j\cdot}}\right)\right\}. \tag{6}$$

In this equation:

- τ_{ij} photoelectric absorption coefficient of radiation of energy i in element j;
- K $(r-1)/r$, where r is the K absorption jump;
- ω K fluorescence yield;
- t_{ij} $\tau_{ij}{}^{\alpha}g_i{}^{\alpha} + \tau_{ij}{}^{\beta}g_i{}^{\beta}$;
- g^{α} Kα transition probability;
- g^{β} Kβ transition probability ($g^{\alpha} + g^{\beta} = 1$);
- μ' μ cosec ϕ;
- μ'' μ cosec ψ;
- $\bar{\mu}_{ij}$ weighted mean mass-absorption coefficient for Kα and Kβ radiation: $\bar{\mu}_{ij} = \mu_{ij}{}^{\alpha}g_i{}^{\alpha} + \mu_{ij}{}^{\beta}g_i{}^{\beta}$.

The parameter c_{4321} expresses the tertiary excitation of element 1 by energy 4 via element 3 and via element 2. It is the solution of a fivefold integral:

$$c_{4321} = \frac{\tau_{43}K_3\omega_3 t_{32}K_2\omega_2 t_{21}}{4\tau_{41}}T(\mu'_{4\cdot}, \mu''_{1\cdot}, \bar{\mu}_{3\cdot}, \bar{\mu}_{2\cdot}). \tag{7}$$

J. Sherman [2] published an analytical expression for T in 1959, but it contains a poorly converging series. In 1971 G. Pollai and H. Ebel [5] published a numerical expression for T, which, although it contains a single integral, can quickly be calculated on a computer. It is evident that equations (2) and (3) apply to the secondary and tertiary excitation only if the energy of the K radiation of the exciting element is higher than that of the absorption edge of the element to be excited. If this is not the case, $c = 0$.

The situation is more complicated if the absorption edge of the element to be excited lies in between the Kα and the Kβ lines, because in that case the excitation is caused only by the Kβ radiation. This results in: $t_{ij} = \tau_{ij}{}^{\beta}g_i{}^{\beta}$ and $\bar{\mu}_{ij} = \mu_{ij}{}^{\beta}$. Due allowance for all these discontinuities has been made in our computer programs.

Of course, the situation becomes even more complicated when the excitation is not mono-energetic but poly-energetic. If the exciting source has a spectral energy distribution given by $I(E)$, we can write for the relative intensity v_j:

$$v_j = \frac{w_j \int_{E_{Kj}}^{E_{\max}} \frac{\tau_{Ej} I(E) B}{\mu_{E\cdot} + \mu_{j\cdot}A}\mathrm{d}E}{\int_{E_{Kj}}^{E_{\max}} \frac{\tau_{Ej} I(E)}{\mu_{Ej} + \mu_{jj}A}\mathrm{d}E}, \tag{8}$$

where:

- τ_{Ej} = photoelectric absorption coefficient of radiation of energy E in element j;
- μ_{Ej} = mass absorption coefficient of radiation of energy E in element j;
- B = term in parentheses from equations (2) and (3); without secondary and tertiary excitation it is equal to 1;
- E_{Kj} = energy of K absorption edge of element j.

We apply this equation to the case of bremsstrahlung spectra of β emitters, such as ^{147}Pm, where $I(E)$ is determined experimentally.

This integral presents problems for X-ray tube spectra since the continuous bremsstrahlung spectrum has superimposed on it the characteristic K or L lines (or both) of the anode material (for example, Cr, W, Mo), which make a significant contribution but are difficult to integrate numerically. J. V. Gilfrich and L. S. Birks [6] have carefully measured the spectra for various types of tube in the wavelength range from 0.027 to 0.299 nm (46-4.15 keV) using a wavelength interval of 0.002 nm. Equation (8) can be modified to suit these measurements; it then becomes:

$$v_j = \frac{w_j \sum_{i=0}^{i_{\max}} \frac{\tau_{ij} I_i B}{\mu_{i\cdot} + \mu_{j\cdot}A}}{\sum_{i=0}^{i_{\max}} \frac{\tau_{ij} I_i}{\mu_{ij} + \mu_{jj}A}}, \tag{9}$$

where I_i is the intensity of the exciting radiation in the wavelength interval i, with the zeroth interval beginning at 0.027 nm.

The index $i_{\max}$ relates to the wavelength interval just below that of E_{Kj}. The tables end at $i = 136$ ($\lambda = 0.299$ nm). In the computer calculations we use some tabulated values and some values given in the form of empirical relations for the parameters $\mu(Z,\lambda)$, $\tau(Z,\lambda)$, $\omega(Z)$, $g^{\alpha}(Z)$ and $r(Z)$. In this way all the parameters are taken from the memory, or calculated if the atomic numbers are given to the computer.

Summary. X-ray fluorescence analysis is based on the emission of a characteristic X-ray spectrum when an element is irradiated by hard X-rays, gamma rays or fast electrons. Formerly the relation between the concentration of the element and the intensity of the excited characteristic radiation was determined entirely by means of reference samples, but in recent years it has become possible, under certain conditions, to find this relation purely by calculation. We have experimentally tested the reliability of the method used, called the fundamental-parameter method. We have also used the method for calculating favourable experimental conditions for analyses using reference samples. The present article describes the calculation and experimental verification of the calibration curves for two binary systems (Ag-Sn and Fe-Ni, the first being excited with mono-energetic radiation and the second with poly-energetic radiation) and for a ternary system (Fe-Co-Ni). The relative intensity is calculated as a function of the excitation energy for a Cu-Ag alloy, which enables us to calculate the excitation energy required to obtain a suitable calibration curve.

The ratio of the intensity of the CrKα line excited by tertiary radiation to that excited by primary radiation from a number of Cr-Fe-Ni alloys is also calculated. It is shown that the contribution from the tertiary intensity can be considerable. Finally, there is a calculation of the relative intensities as a function of the element concentrations in the system Cs-Br-H_2O, showing that the effect of the dilution with water is compensated to a large extent by the lower absorption of water.

[5] G. Pollai and H. Ebel, Spectrochim. Acta 26B, 761, 1971.
[6] J. V. Gilfrich and L. S. Birks, Anal. Chem. 40, 1077, 1968.

Philips tech. Rev. 34, 344-349, 1974, No. 11/12

The determination of carbon, oxygen and nitrogen in semiconductors by spark-source mass spectrography

J. B. Clegg and E. J. Millett

Introduction

The electrical properties of the group-IV elements silicon and germanium are primarily determined by the free-carrier concentration, which is simply related to the concentrations of the singly ionized donor and acceptor elements of group V and group III always present in small quantities in silicon and germanium. Other chemical elements either have more complex effects (introducing trapping centres, reducing free-carrier mobility or minority-carrier lifetimes, forming complexes with vacancies or with other impurity elements, etc.) or they may be inactive or virtually absent from the lattice. Most of these effects have been unravelled over the last twenty years by patient experiment as the techniques of purification and analysis developed together. One of the difficulties in this work has been that no single analytical technique could determine all the impurity elements with the required sensitivity at once, and that some elements, particularly carbon, oxygen and nitrogen are difficult to determine by any chemical method.

In the III/V semiconducting compounds, the interpretation of the free-carrier concentration is complicated by the fact that the group IV elements behave as both donors and acceptors, and oxygen and nitrogen may play more significant roles, determining, for example, the emission wavelength of light-emitting diodes. Many of the analytical techniques developed for silicon and germanium are either inapplicable to the III/V compounds or would require extensive re-development for each host lattice in turn.

In principle the mass spectrograph with spark source could offer a solution here. With this instrument nearly all the elements can be determined at once in a single analysis, at a sensitivity that is virtually unaffected by the nature of the host lattice [1]. For our purpose, however, the conventional equipment cannot be used, since the results it gives for C, O and N are subject to serious errors, and may therefore be completely misleading. The interferences come mainly from two sources: one is the residual gas, and the other is the surface of the sample being investigated. By making certain modifications to equipment and method we have been able to reduce the effect of these sources of interference sufficiently to permit the concentration of carbon, oxygen and nitrogen to be measured down to values of only 10^{-2} ppma (parts per million atomic).

J. B. Clegg, B.Sc., and E. J. Millett, M.Sc., are with Mullard Research Laboratories, Redhill, Surrey, England.

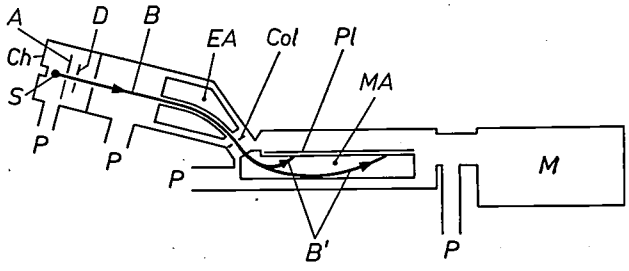


Fig. 1. Schematic cross-section of the spark-source mass spectrograph. *Ch* chamber with ion source. *S* electrodes of sample material. *B* ion beam. *A* accelerating electrodes with slit. *D* beam-defining slit. *EA* electrostatic analyser. *Col* collector plate. *MA* magnetic analyser. *B'* mass-resolved ion beams. *Pl* photographic plate. *M* magazine for the photographic plates. *P* pump connections.

A vacuum discharge maintained between the electrodes *S* provides a source of ions representative of the sample composition. These ions are accelerated from the ionization region and, after passing through the electrostatic analyser (energy selector), are brought to a line focus on an ion-sensitive photographic plate mounted within the magnetic analyser (mass selector). High resolution is achieved by means of the double-focusing configuration of both the electrostatic and magnetic analysers. The number of ions reaching the photographic plate (the 'exposure') is derived by measuring the total quantity of charge incident on the collector plate, which intercepts 50% of the unresolved ion beam. Such a spectrograph is usually pumped with oil-diffusion pumps fitted with liquid-nitrogen cold traps; in the analysis regions the pressure is about 0.003 mPa (2×10^{-8} torr), but only about 0.3 mPa (2×10^{-6} torr) in the ion-source chamber. This is far too high for an accurate determination of traces of C, O and N.

Using the conventional spark-source mass spectrograph (*fig. 1*) two rod-like electrodes of the sample material (approximately $1\times3\times10$ mm) are partially evaporated and ionized by a pulsed radio-frequency spark, and a mass spectrum is recorded on a photographic plate. On the developed plate the impurities can be identified by eye, provided the isotope pattern for each element is known. The approximate concentrations can be deduced by visual comparison with the less abundant isotopes of the elements forming the matrix, or with an impurity of known concentration. For accurate quantitative work the individual line profiles have to be measured with an optical microdensitometer, and the measured optical density converted to ion intensity. The conversion is relatively

complex and the procedures we use are based on extensive experimental work carried out using an automated microdensitometer coupled to a computer [2].

The efficiencies of the processes of ionization, transmission through the spectrograph and image formation are not identical for each element so that the sensitivities of the elements are not exactly the same. Absolute accuracy can then be achieved only by calibration, as with any other instrumental technique, using previously analysed standard samples. Accuracies of 5-10% can then be attained [3]. The provision of reliable standard samples presents special problems, as they must be genuinely homogeneous over the volume sampled (of the order of a cubic millimetre) and ideally there has to be an alternative analytical technique with a sensitivity sufficient to provide reliable data at concentrations close to the limit of detection of the mass spectrograph. In comparing analysed standard samples, the experimental conditions must be rigorously controlled or serious errors may be introduced. Confusion between errors of measurement, errors of standardization, interference and experimental irreproducibility has often thrown doubt on the quantitative reliability of the spark-source mass spectrograph. The stable experimental conditions and reliable procedures used in our work played an important part in establishing the conditions required to determine the concentrations of these difficult elements, C, O and N.

We first describe the various sources of interference and their elimination and then discuss the results of comparisons with standard samples analysed by specific optical absorption and activation techniques which confirm the quantitative reliability of the method.

The residual gas as a source of interference

The first source of interference originates in the residual gas, since the spark discharge does not take place in an absolute vacuum. During the discharge, molecules of residual gases such as CO and H_2O are subject to local dissociation and ionization. Some of the ions produced are accelerated, leave the source and travel into the analyser regions of the spectrograph; these ions then contribute to the carbon and oxygen line intensities at the photographic plate. We have studied this process quantitatively by deliberately introducing pure gases into the ion-source chamber during the analysis of a gallium-arsenide sample. *Fig. 2* shows the results obtained with CO. It can be seen that the CO molecule is ionized directly to form CO^+ and also dissociated and ionized to form equal numbers of C^+ and O^+ ions. In both cases, the concentration of the ions produced is linearly dependent on the gas pressure.

In the standard version of the mass spectrograph this interference leads to serious errors, since many of the sample ions bombard the chamber walls during the analysis and liberate large quantities of CO and other gases. This unwanted gas load raises the working pressure in the chamber to about 0.3 mPa (about 2×10^{-6} torr). At this pressure the carbon and oxygen from the residual CO give rise to interferences equivalent to a concentration of 1 ppma. Extrapolating the curves of fig. 2, we see that to maintain an acceptable background level, e.g. 0.01 ppma, the pressure in the ion-source chamber should not be greater than about

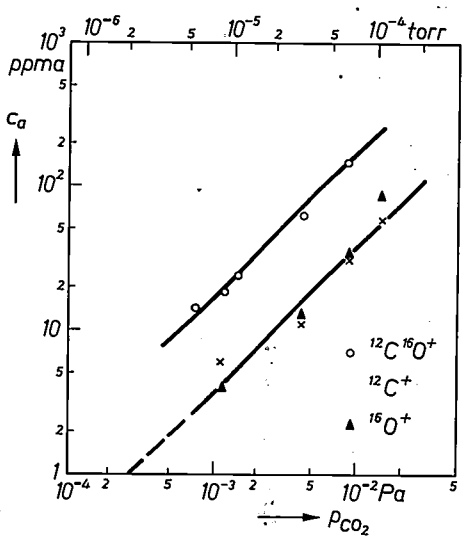


Fig. 2. The apparent concentrations of the ions CO^+, C^+ and O^+ in gallium arsenide as a function of the pressure of carbon monoxide in the ion-source chamber. Carbon monoxide is ionized directly to produce CO^+ ions and also dissociated and ionized to give equal numbers of C^+ and O^+ ions. At a gas pressure of 3×10^{-4} Pa, the estimated concentrations of both C^+ and O^+ are about 1 ppma. The electrodes used in this experiment contained about 0.5 ppma of carbon and oxygen.

3 μPa (2×10^{-8} torr). With the restricted size of the chamber it is not possible to achieve the desired pumping speed with conventional pumps, mainly because of the resistance of the pipe-work connecting the pump to the chamber. We have therefore connected the chamber directly to a cryogenic pump with its pumping surface cooled to the temperature of liquid helium.

Fig. 3 shows a cross-section of the equipment. A thin-walled stainless-steel vessel forms the reservoir for the liquid helium and the base acts as the principal conden-

[1] A. J. Ahearn, F. A. Trumbore, C. J. Frosch, C. L. Luke and D. L. Malm, Anal. Chem. 39, 350, 1967.
A. J. Ahearn *(ed.)*, Mass spectrometric analysis of solids, Elsevier, Amsterdam 1966.
A. J. Ahearn *(ed.)*, Trace analysis by mass spectrometry, Academic Press, New York 1972.
[2] E. J. Millett, J. A. Morice and J. B. Clegg, Int. J. Mass Spectrom. Ion Phys. 13, 1, 1974.
[3] R. K. Skogerboe, A. T. Kashuba and G. H. Morrison, Anal. Chem. 40, 1096, 1968.

sation or pumping surface. The sides of the vessel are screened from room-temperature radiation by a cylindrical radiation shield cooled by liquid nitrogen, and the base is screened by an array of cooled chevron baffles. Normally these baffles are designed so that they are 'optically tight', but in our case this would have substantially reduced the pumping speed in the chamber. We therefore modified the geometry of the baffles to expose 50% of the pumping surface to the ion-source chamber, to obtain increased pumping speed at the expense of higher helium consumption.

The pumping speed of this system is 0.4 m^3/s for nitrogen gas at 15 μPa (10^{-7} torr), enabling low pressure to be attained rapidly (see *fig. 4*). A lowest pressure of 0.05 μPa (4×10^{-10} torr) is obtained within 40 minutes of the admission of liquid helium; this pressure is almost two orders of magnitude lower than that attainable with the standard diffusion-pump system. Pressures measured with the spark running are typically about 3 μPa (2×10^{-8} torr), again representing two orders of magnitude of improvement.

The rapid pump-down time makes it possible to analyse two samples per day, an important consideration in the routine analysis of high-purity materials. One filling of the helium cryostat lasts approximately 4 hours, sufficient time to carry out a complete analysis.

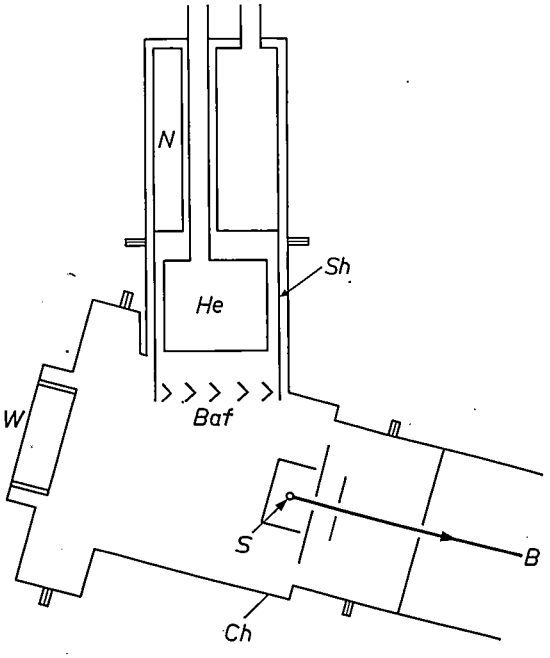


Fig. 3. Schematic diagram of helium cryogenic pump mounted on the ion-source unit, as used in our mass spectrograph for trace analysis. *W* glass window. *He* vessel containing 500 ml of liquid helium. *N* vessel containing liquid nitrogen. *Sh* radiation shield. *Baf* cooled chevron baffles. The other symbols have the same significance as in fig. 1. The lower surface of the helium vessel, which acts as the principal pumping surface, is polished to a mirror finish and is shielded from room-temperature radiation by an open-structured chevron-baffle system cooled to liquid-nitrogen temperature. One fill of the helium vessel lasts 4 hours.

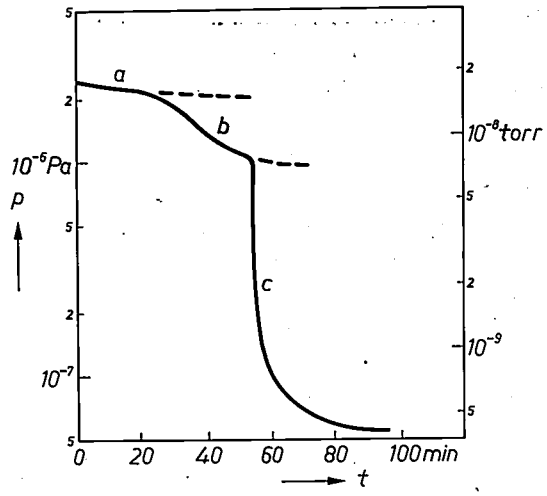


Fig. 4. Pressure variations in the ion-source chamber for various methods of pumping. Curve *a*: diffusion pump only; curve *b*: in addition, cryogenic pumping with liquid nitrogen; curve *c*: cryogenic pumping with liquid helium. Cryogenic pumping with liquid nitrogen reduces the ultimate pressure to 1 μPa by the condensation of water vapour and certain hydrocarbons, but effective pumping for other gases is achieved only when the condensation surface is cooled to the temperature of liquid helium. These pressure measurements were made without the radio-frequency spark running and after the pump and the ion-source chamber had been baked at 150 °C for 12 hours and allowed to cool to room temperature.

Interference from surface contamination

It is standard practice to clean semiconductor surfaces by a combination of mechanical polishing, chemical etching and treatment with organic solvents. These procedures are effective in removing a broad range of impurities, but unfortunately they often introduce significant quantities of carbon and oxygen. For example, germanium is normally etched in a mixture of nitric, acetic and hydrofluoric acids to produce what is thought to be a clean surface. Analysis of the surface showed however that there are about 3×10^{15} atoms of oxygen present per cm^2, corresponding to a surface coverage of about 3 monolayers [4].

A more effective way of producing a clean surface in the mass spectrograph is the *in-situ* pre-sparking of the sample electrodes, prior to recording the analytical exposures. This process depends on the radio-frequency spark evaporating small quantities (about 1 to 10 mg) of surface material to expose the uncontaminated surface.

The effectiveness of this cleaning procedure is illustrated by an analysis of high-purity indium phosphide. *Fig. 5* shows the concentrations of carbon and oxygen as a function of depth eroded into the electrodes. The initial concentrations, due to the presence of these elements at the surface, decay rapidly to reach essentially constant levels. As the carbon and oxygen content of this sample is known to be low [5], it is highly probable that these final levels (fig. 5) are indicative of the con-

centrations present *in* the sample. From these results it is apparent that the removal of about 140 microns of material reduces the surface interferences to about 0.01 ppma or lower. Without this cleaning procedure, erroneous concentrations in the region of 0.5 ppma are observed.

This procedure will not be effective however in the presence of significant quantities of oxygen or water, as a partial pressure of only 0.1 mPa (10^{-6} torr) is sufficient to produce one monolayer of oxide in a second. This level of contamination would represent more than 1 ppma in the evaporated sample. With the reduced partial pressures obtained using the cryogenic pump, however, this source of interference is maintained at a low level.

It should also be noted that when mass-spectrograph measurements are compared with results obtained from activation analysis — which amounts to the measurement of a gamma spectrum — exact agreement is not expected, since fractions of a ppm are involved. The values given by the mass spectrograph can differ from the true values by a factor of three in either direction. This is due to various effects, such as differences in volatility and ionization potential for the various elements, instrumental errors, etc. If standard samples are available an empirical correction factor can be determined, the *relative sensitivity coefficient*, which can be applied to subsequent measurements.

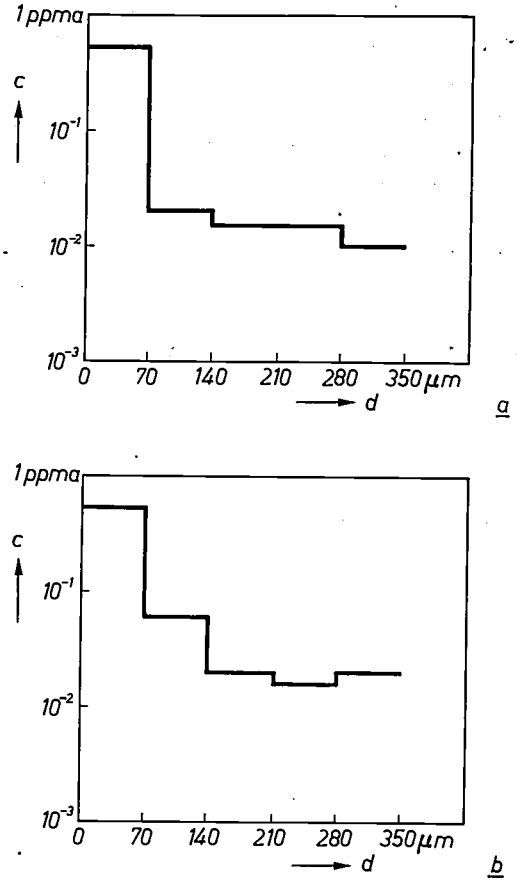


Fig. 5. The measured concentration c of oxygen (a) and carbon (b) as a function of depth eroded into indium phosphide. (The depths eroded are derived from the amount of electrode material consumed during each exposure.) The concentrations found in the first 70 microns are most probably due to surface contamination. After the evaporation of 140 microns, the concentrations of both impurities are reduced by at least a factor of 25 to produce an effectively clean surface.

Analytical performance

In the previous sections, we showed how the level of the interference can be reduced by using an improved method of evacuating the ion-source chamber, combined with *in-situ* cleaning of the electrodes. The effect of this improved analytical technique on quantitative response is illustrated by the analysis of three semiconductor materials, germanium, silicon and indium phosphide.

Table I shows the analytical results obtained from a sample of vacuum-grown germanium. Two analyses were carried out, one using the standard pumping system and the other using the cryogenic pump. It can be seen that the improved pumping leads to a significant reduction in the concentrations of C, O and N. The values measured when the standard pumping system was used (C = 2 ppma, O = 1 ppma), are entirely due to the residual gas. Analysis of this germanium sample by both vacuum-fusion analysis and infrared spectrophotometry [4] showed that the oxygen concentration was lower than the detection limit of either method, 0.5 ppma and 0.1 ppma respectively. The observed mass-spectrographic value (0.05 ppma) is thus consistent with these results.

Table I. Analysis of vacuum-grown germanium using two types of ion-source pumping. For both analyses the samples were given the same *in situ* cleaning by pre-sparking. The standard deviations (66% confidence limits) are given for both carbon and oxygen [6]. (Analysis of this material by vacuum fusion and infrared spectrophotometry gave oxygen concentrations of < 0.5 and < 0.1 ppma respectively.

Type of pumping	Concentration ppma		
	Carbon	Oxygen	Nitrogen
Diffusion	2	1	0.1
Helium cryogenic	0.07 ± 0.01	0.05 ± 0.02	< 0.04

[4] E. J. Millett, L. S. Wood and G. Bew, Brit. J. appl. Phys. **16**, 1593, 1965.

[5] Independent analyses using γ-photon activation have given the following concentrations: carbon about 0.06 ppma, oxygen < 0.3 ppma. Thanks are extended to J. Hislop, A.E.R.E., Harwell, for carrying out these analyses.

[6] All the spectrographic results in this article have been corrected for the known dependence of photographic sensitivity on ion mass. See for example E. B. Owens and N. A. Giardino, Anal. Chem. **35**, 1172, 1963.

It is frequently the case that the concentrations of C, O and N present in semiconductors are below the limit of detection of other independent analytical techniques. Confirmatory analyses, to establish the reliability and accuracy of this method, are therefore not generally possible. One way round this problem is to deliberately dope semiconductors with C, O and N so that the higher concentration levels may be measured reliably by other techniques. With germanium, the oxygen content has been increased either by adding GeO_2 to the crystal melt during growth or by growing the crystal under an atmosphere of oxygen [7]. Analysis of two oxygen-doped samples by infrared spectrophotometry gave concentrations of 1.0 and 4.6 ppma, while the uncalibrated mass-spectrographic concentrations were 3.3 and 11 ppma respectively. The agreement between the results is satisfactory, bearing in mind that the spectrographic values 3.3 and 11 are as yet uncorrected for relative sensitivity.

Such measurements define the relative sensitivity coefficient for oxygen in germanium. There is no fundamental reason why this calibration should not apply to the determination of much lower concentrations, provided that the necessary precautions are taken to reduce the sources of interference. Referring to the results of Table I, it follows that the corrected oxygen concentration in germanium is about 0.02 ppma, a value which cannot be attained by other routine analytical methods.

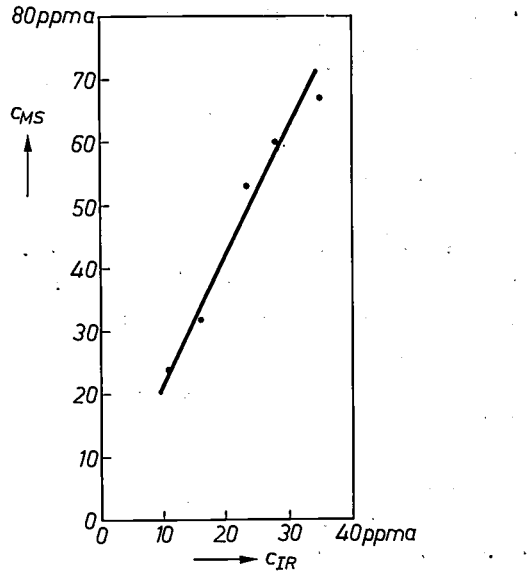


Fig. 6. Comparison of the oxygen concentration (in Czochralski-grown silicon) measured by infrared absorption (*IR*) and by mass spectrography (*MS*). The proportional relationship shows that oxygen is quantitatively measured by the mass-spectrographic technique and that the oxygen has a relative-sensitivity factor of 2.1 in the silicon host lattice.

Table II. Analysis of indium phosphide by mass spectrography and γ-photon activation analysis. Where a sufficient number of individual measurements were made (i.e. more than three), the standard deviation is given. Two of the samples were undoped while the third (L449) was doped with oxygen by the addition of indium oxide to the melt during growth.

Crystal type	Concentration ppma			
	Carbon		Oxygen	
	Mass spectrography	γ-photon	Mass spectrography	γ-photon
L170 undoped	0.12 ± 0.03	0.1	0.07 ± 0.01	0.4
L374 undoped	0.04 ± 0.03	0.06 ± 0.05	0.02 ± 0.01	< 0.3
L449 oxygen-doped	0.07 ± 0.04	< 0.06	0.18 ± 0.02	0.8 ± 0.3

Silicon grown by the Czochralski technique is always unintentionally contaminated with oxygen originating from the silica (SiO_2) crucible. Analysis of this silicon by both mass spectrography and infrared spectrophotometry [8] gives results which show a good linear correlation with each other (see *fig. 6*). This shows that the mass-spectrographic method gives a quantitative measure of the oxygen content over the concentration, at least in the range measured (10-40 ppma). Here again we can extrapolate this calibration to make absolute measurements at concentrations which are difficult to measure by the infrared technique. For example, oxygen contents of about 0.1 ppma have been determined for small samples of silicon grown by the floating-zone technique.

Results on several indium-phosphide [9] samples, obtained by both mass-spectrographic analysis and γ-photon activation analysis, are shown in *Table II*. With carbon, good agreement is obtained between the two techniques. With oxygen the agreement is not so good, but for the two sets of results (L170 and L449) where a direct comparison is possible the relative-sensitivity coefficient is essentially constant. It is interesting to note that an unmodified spectrograph in another laboratory gave carbon and oxygen concentrations of 2 and 0.8 ppma respectively for sample L170.

[7] J. R. Dale and J. C. Brice, Solid-State Electronics **3**, 105, 1961.
[8] J. A. Baker, Solid-State Electronics **13**, 1431, 1970.
[9] Thanks are extended to J. B. Mullin, Royal Radar Establishment, Malvern, for supplying the indium-phosphide samples.

These results demonstrate the large errors which occur if the sources of interference are not sufficiently reduced.

Finally, we should make a few comments about the reproducibility of the results. With the oxygen-doped indium-phosphide sample (L449), where the oxygen line at the photographic plate is intense and easily measured, the relative standard deviation was 10%. Considering the low levels of concentration involved, this result shows both the excellent reproducibility of the method and the high degree of homogeneity of the doped sample. (Only about 0.3 mm^3 of material is consumed during each exposure and any compositional changes between successive volumes would lead to scatter in the experimental measurements.)

In conclusion, we may say that this work has shown that the spark-source mass spectrograph can be used with confidence for the determination of C, O and N down to 10^{-2} ppma. This extension of the mass-spectrographic technique provides for the first time the possibility of determining those elements that have previously fallen outside the scope of routine analytical methods. It thus provides impurity information on any semiconductor, covering essentially the whole periodic table.

Summary. By connecting the ion-source chamber of a spark-source mass spectrograph directly to a liquid-helium cryogenic pump it has been possible to eliminate the serious interferences that can originate from the C, O and N in the residual gas. Careful 'pre-sparking' in the source at low pressure (about 1 μPa) removes the oxide layer and the carbon from the surface, both for semiconducting compounds and for pure elements; this was a second important source of interference. The use of these two techniques in combination with a good measurement procedure enables analytical results to be obtained that agree well with those obtained in other ways. The elements C, O and N can now be determined with confidence down to concentrations of 10^{-2} ppm (atomic).

Philips tech. Rev. 34, 350-355, 1974, No. 11/12

Stoichiometric analysis of gallium selenide

E. Bruninx and L. C. Bastings

Introduction

By stoichiometric analysis of a compound we mean the accurate quantitative determination of its constituent elements.

According to the classical laws of chemical combination (Dalton) the composition of a compound can be represented with simple integral numbers. This is the 'stoichiometric' composition. Modern theories show however that in the solid phase the composition of a compound varies within the limits of the 'existence region' [1], which is connected with the existence of point defects in the crystal lattice. Insulators and semiconductors have a narrow existence region; an example is CdTe, in which the existence region is as narrow as 10^{-3} at%. A wider existence region is found in intermetallic compounds; in CuAu, for example, the width of the existence region is 1 at%.

In semiconducting compounds differences in composition even within the narrow limits mentioned can have a very marked effect on the physical properties, and the investigation and reproducible preparation of these substances therefore require a very accurate knowledge of their composition. It is also very important to distinguish these effects from those due to traces of other elements. This means that the analytical chemist must be able to determine the major constituents of a compound with a very high degree of accuracy.

To obtain the accuracy required, classical chemical methods, in particular gravimetry, titrimetry or coulometry, have generally been used in the past for carrying out stoichiometric analyses. The high accuracy of these methods is essentially due to the high accuracy with which substances can be weighed. A difficulty with classical chemical methods, however, is that they are not strictly specific for a given element. This means that an analysis of a composite substance requires a detailed study of undesirable interelement effects.

The accuracy of physical methods of analysis has in the past been regarded as inferior to that of classical chemical methods. We have made a detailed investigation of possible sources of error to see whether one of these physical methods could perhaps be improved to give results equivalent to those obtainable with the classical chemical methods. The instrumental method that seemed to offer the most chance of success was neutron activation analysis, mainly because the interelement effects with this method are weakest. The investigation was made with gallium selenide.

In the analysis of gallium selenide by neutron activation analysis we were able to reduce the total uncertainty to approximately 0.1%. This is a considerable improvement compared with the uncertainty usually typical of activation analyses (2%), but the lowest uncertainty that can be reached with classical methods (less than 0.01%) could not be achieved, and the investigation showed that this would not in fact be a practical proposition.

The requirement of extreme accuracy

To achieve the ultimate in accuracy which a given method of measurement is capable of providing, it is necessary to investigate the extent to which every single step of the measurement process contributes to the uncertainty in the final result, taking random and systematic uncertainties into account separately, of course (*fig. 1*). Efforts should next be made to reduce the main contributions to the random uncertainty and to gain a thorough quantitative knowledge of the systematic uncertainties, if possible eliminating them. They can often be eliminated by analysing reference samples under identical conditions, which really amounts to making an 'automatic' correction. In attempting to improve the accuracy of an analytical method it is of course important to remember that for the purposes of a stoichiometric determination there is no point in trying to achieve a much higher accuracy than that to which the atomic weights of the elements in question are known. The random uncertainty [2] given with the analytical results in this article is the standard deviation.

The standard deviation s of a series of n observations $x_i (i = 1, \ldots, n)$ is given by the expression

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}},$$

where $\bar{x}$ is the mean value of x_i. The standard deviation of $\bar{x}$ is $s/\sqrt{n}$. In a composite measurement the standard deviation of the whole is found by summing the squares of the standard deviations of the parts:

$$s_{tot}^2 = s_1^2 + s_2^2 + \ldots .$$

(Systematic deviations must of course be summed linearly.)

Dr E. Bruninx and L. C. Bastings are with Philips Research Laboratories, Eindhoven.

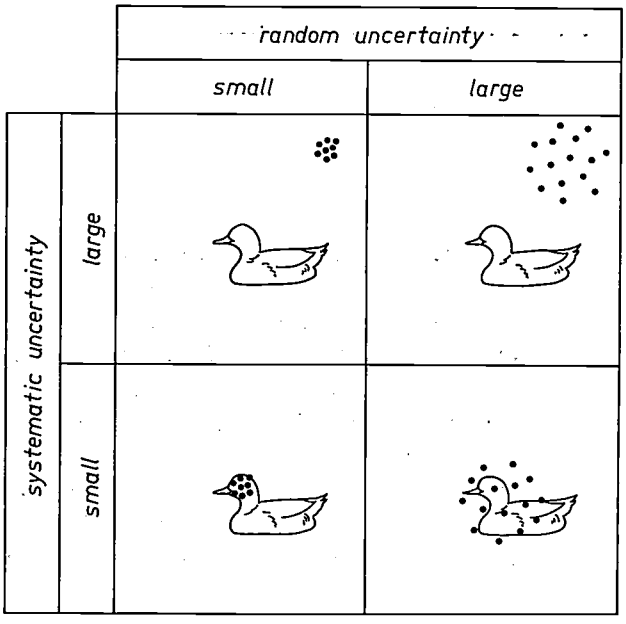


Fig. 1. The nature of the random and systematic deviations that can occur in the results of a measurement, illustrated by shots fired at a target; each shot represents one measurement.

To trace systematic deviations the analytical chemist must compare the results of his measurements with those found in other ways. We therefore compared the measured gallium content with the results of an equally refined titrimetric method, and found good agreement between them.

To have at the same time a comparison with a method by which selenium as well as gallium could be determined, we finally performed very careful analyses using the X-ray fluorescence method: A critical examination of the sources of errors showed that with this method, contrary to original expectations, it should be possible to reduce the total uncertainty to less than 0.1 %. Since an X-ray fluorescence analysis takes much less time, it is evidently more attractive than the other two methods for practical reasons.

We also used a check of an entirely different type, in which we compared the standard deviation in a series of analyses with the value calculated from the standard deviations of the individual steps in the analysis procedure, to see if they agreed.

Neutron activation analysis

In neutron activation analysis by gamma spectrometry the normal procedure [3] is to make a sample radioactive by bombarding it with neutrons and then to detect the gamma radiation emitted by the sample with a Ge(Li) counter, perhaps repeating the measurement several times. The detector gives output pulses whose height is proportional to the energy of the gamma quanta counted. The pulses are then sorted by height, giving a gamma spectrum of the sample. The lines appearing in the spectrum reveal the constituent elements of the sample, and their intensity is a measure of the concentration of the elements.

Our problem is of course very different from that encountered in the analysis of an unknown sample: we are only concerned with two elements, and these are already known. We were able to use a different procedure from the normal analytical method [4], enabling us to achieve the required low uncertainty, since the half-lives of the two nuclides resulting from the capture of thermal neutrons differ very considerably: the half-life of ^{72}Ga is 14.2 hours; that of ^{75}Se is 128 days. This enabled us to determine the radioactivity of each of the elements separately, without the need for pulse-height discrimination. The gallium determination was made after a short irradiation, in which for all practical purposes only gallium was activated. Selenium was determined after prolonged irradiation, followed by a waiting time of two weeks, which was long enough for the complete decay of all the ^{72}Ga produced. As an additional check on the purity of the sample, a careful mathematical analysis was made of both decay curves to make sure that they corresponded to the known half-life of each element.

The samples for analysis and the reference samples were solutions in nitric acid. The use of a solution has the advantage that both analysis and reference samples are completely homogeneous and in the same physical state. Another advantage is that the concentration can easily be varied over a wide range. The required quantities were weighed out with a chemical balance and not measured out volumetrically. With a suitable balance, and provided the quantities of the sample available are not too small, the weights can be obtained with an extremely small uncertainty. The random uncertainty in the weighing of the samples was approximately 10^{-3} %; the systematic uncertainties in this procedure are mainly due to inaccuracies in the weights.

The activity of the irradiated samples was measured with a scintillation counter. The activity A_1 of a sample 1 is given by:

$$A_1 = kM_1(\sigma\Phi)_1, \qquad (1)$$

where M_1 is the number of nuclei of Ga or Se, $(\sigma\Phi)_1$ is the product of the effective activation cross-section σ

[1] See for example W. Albers and C. Haas, Philips tech. Rev. 30, 82, 1969.
[2] We prefer the term 'random uncertainty' to the commonly used terms 'random error' and 'precision'.
[3] The principle of neutron activation analysis is described in the article by M. L. Verheijke in this issue, page 330.
[4] E. Bruninx, Anal. chim. Acta 67, 17, 1973.

and the neutron flux Φ, and k is a proportionality factor. The factor $(\sigma\Phi)_1$ cannot be determined with sufficient accuracy, and varies from one position to another in the nuclear reactor. This difficulty can be avoided by irradiating another sample of known composition (2), for which $\sigma\Phi$ has the same value. The number of nuclei is then given by

$$M_1 = M_2 \frac{A_1(\sigma\Phi)_2}{A_2(\sigma\Phi)_1} = M_2 \frac{A_1}{A_2}, \qquad (2)$$

which can be used for calculating M_1.

The value of σ in (1) and (2) cannot be regarded as a constant, since the probability that a neutron will be captured by an atomic nucleus depends to a fairly great extent on the energy of the neutron. Every irradiation thus yields a mean value that depends on the velocity distribution of the neutrons — and hence on such quantities as the percentage of epithermal neutrons. The effective value of the activation cross-section used in our calculations thus varies from one position to another in a nuclear reactor, and even depends to some extent on the nature of the sample.

To make the factors $\sigma\Phi$ as close in value as possible, our procedure in all cases was to expose a combination of two analysis samples and four reference samples to simultaneous irradiation while continuously changing their positions (*fig. 2*). So as to obtain the maximum accuracy we chose the reference samples in such a way that two of them had a content about 20% below the expected value, one close to it and one about 20% above it.

Experiments with six identical reference samples showed that the standard deviation in $\sigma\Phi$ was reduced in this way to about 0.1%.

We now come to the errors that occur in the measurement of the gamma radiation. Here again there are two types, random and systematic. The random uncertainty s in a count of gamma quanta includes a contribution s_p arising because the gamma radiation consists of discrete particles and a contribution s_i from the electronic instrumentation: $s^2 = s_p^2 + s_i^2$. Statistical theory shows that s_p is equal to the square root of the number of pulses counted. This means that if we want to achieve an uncertainty of about 0.1% we must count at least 10^6 pulses and s_i must be of the same order or smaller; the latter requirement can be met with existing measuring instruments. *Fig. 3* shows how the relative standard deviation of the counting instruments used in all our experiments depends on the number of pulses counted. As can be seen, the smallest random uncertainty obtainable is in the region of 0.05%.

A major *systematic* error can be caused by differences in the level to which the sample fills the polythene cylinder used in the equipment as the sample holder. If the sample holder is placed as usual in a cavity at the top of the scintillation crystal (*fig. 4*), a small difference in level causes a substantial difference in detection efficiency. We were able to reduce this difference considerably by drilling a hole right through the crystal and placing the cylinder in the middle of the crystal. This has the added advantage that the detection efficiency itself is then higher than in the conventional situation.

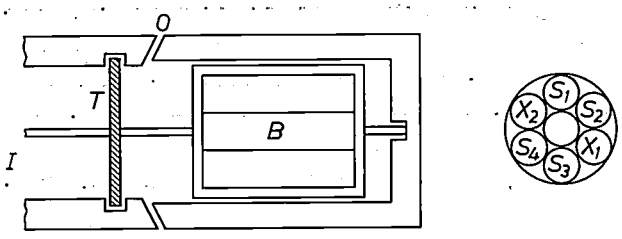


Fig. 2. Schematic cross-section of the sample holder used during irradiations in a nuclear reactor for stoichiometric analysis by neutron activation. *T* turbine, *I* air inlet for driving the turbine. *O* outlet. *B* magazine containing the six samples (two analysis samples *X* and four reference samples *S*). *Right:* illustrating the positioning of the samples in *B*.

We had to make a further correction for the systematic deviation due to the 'dead time' of the equipment. At high counting rates we have found that it is in fact necessary to take two dead times into account, but this will not be dealt with further here. To calculate the true counting rate N_r from the measured counting rate N_0 we therefore used the equation [5]:

$$N_0 = \frac{N_r}{\exp(N_r\tau) + N_r(\Theta - \tau)},$$

where τ and Θ are the two dead times. This equation corrects the loss of pulses better than the classical equation, which only allows for one dead time. In these corrections the total uncertainty for the dead time is less than 0.1% of the total measured activity [6].

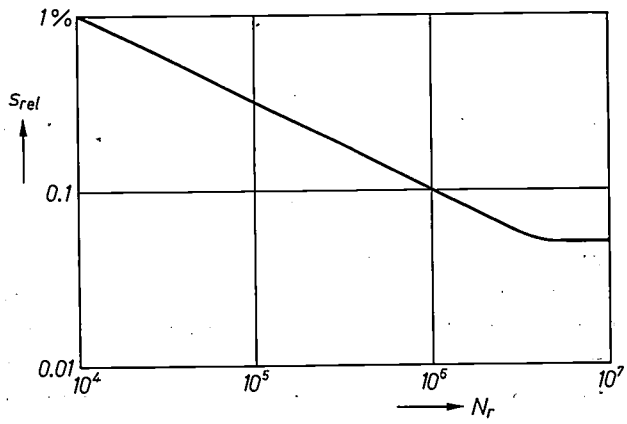


Fig. 3. Relative standard deviation s_{rel} of the number of gamma quanta counted as a function of this number (N). The lower limit is due to the counting instrument.

After every irradiation we had two values of A_1 and four values of M_2 and A_2. Fom these four pairs of values for M_2 and A_2 we determined the proportionality factor between M and A by the method of least squares, and then used this factor to calculate M_1 values from the A_1 values, with the associated standard deviation.

The measures described above enabled us to limit the uncertainty in the final result of the analysis to approximately 0.1 %. This is a significant improvement, because the uncertainty has never previously been less than 2 %. The major contribution to the uncertainty comes from the ratio $(\sigma\Phi)_1/(\sigma\Phi)_2$.

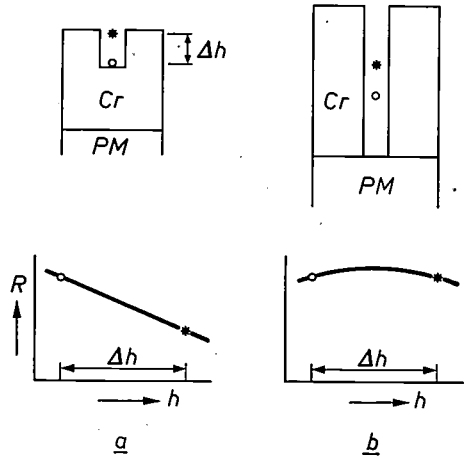


Fig. 4. The cavity in the scintillation crystal *Cr* in which the radioactive sample is placed, in the conventional form (*a*) and in the form we used (*b*). *PM* photomultiplier tube. The graphs show the variation of the detection efficiency R for a thin layer in the cylindrical holder as a function of the position h of the layer.

X-ray fluorescence analysis

As mentioned above, we compared the results of the neutron activation analysis with those obtained by a different physical method, X-ray fluorescence analysis, with the object of tracking down possible systematic deviations in the method of activation analysis.

All samples were analysed in the form of solutions, for the same reasons as in neutron activation analysis. The measurement was based on the counting of the characteristic X-ray fluorescence quanta [7].

The relation between the concentration of the element and the measured X-ray intensity has basically the same form as given by equation (1), with the difference that the factor $(\sigma\Phi)$ is replaced by a factor describing the X-ray excitation conditions and the absorption. It should be noted here that the energy of the X-rays is considerably lower than that of the gamma quanta. Because of the associated strong absorption of the X-rays in the solution, the total quantity of the elements in the analysis and reference samples must be the same.

Owing to the strong absorption of the X-rays the calibration curve is not always a straight line through the origin; it may also be given by:

$$A = kM + B$$

or

$$A = kM + B + CM^2,$$

where B and C are constants.

Using statistical methods we made a choice from among the three equations to establish the shape of the calibration curve. We were able to show that the total uncertainty in analyses of this type can be reduced to less than 0.1 %.

Titrimetric gallium determinations

To check the instrumental analysis we determined the gallium content of the gallium selenide by a classical chemical method. The method we chose was a titrimetric method based on the formation of stable complexes soluble in water.

The complexes involved are compounds of metal ions with other ions, in which the metal ions are completely or partly surrounded and there are coordinate bonds between the surrounding ions and the metal ion [8].

In general, the titrimetric method is less accurate than gravimetry and coulometry. However, by substituting weighing for the volumetric measurement, and by determining the end-point of the titration instrumentally, we made the method described here just as accurate as a gravimetric determination of a simple compound. In gravimetric analyses of samples weighing about 1 gram the error depends on the accuracy of the weighing, which is about 10^{-3} %, as mentioned above.

In a complexometric analysis the complexing agent used (the titrant) must meet certain requirements. In the first place the reaction must be virtually complete. If reference samples are not used, the reaction constant must be accurately known to permit correction for an incomplete reaction [9]. The reaction should also take place fast enough to enable a sharp 'end-point' of the titration to be found. Obviously, a sensitive method of end-point detection must be available.

---

[5] J. W. Müller, Nucl. Instr. Meth. **112**, 47, 1973.
[6] E. Bruninx and H. J. Prins, Int. J. appl. Rad. Isot. **25**, 483, 1974.
[7] The principle of X-ray fluorescence analysis is described in the article by M. L. Verheijke and A. W. Witmer in this issue, page 339.
[8] See F. Basolo and R. C. Johnson, Coordination chemistry, Benjamin, New York 1964.
[9] See J. Kragten, Talanta **18**, 311, 1971.

In the determination of many metals these requirements are very satisfactorily met by the complex-former ethylene diamine tetra-acetic acid (EDTA). We therefore adopted this compound as the titrant for our determination of gallium in gallium selenide, which can be dissolved in nitric acid.

For determining the end-point of the titration instrumentally we used the reaction of EDTA with the gallium complex of methyl thymol blue (MTB). This complex is much less stable than Ga-EDTA, and MTB is therefore easily displaced by EDTA. In solutions this brings about a colour change. We measured the excitation at one wavelength, i.e. the wavelength at which the extinction change is greatest upon the addition of EDTA (*fig. 5*). At this wavelength (600 nm) the transition to constant absorption is most distinct (*fig. 6a*).

The rate of the reaction of EDTA with the gallium-MTB complex is not very fast. This makes the end-point detection rather difficult. In a photometric end-point determination it is not very easy to get around this difficulty by measures such as titration at higher temperature or very slow titration. We therefore performed the titration of the gallium indirectly as follows. A carefully weighed quantity of EDTA solution of known strength was added to the gallium-selenide solution to give a small approximately known excess. This excess of EDTA was determined by using a zinc solution as the titrant; the zinc-EDTA reaction is very fast. The end-point was again determined by photometrically detecting the colour change of methyl thymol blue. The situation differs from that shown in fig. 6*a*: the absorption is now initially constant, and then shows a marked linear increase (fig. 6*b*). The end-point was taken by extrapolation to be the point of intersection between the two straight parts of the curve. In determining the end-point it is necessary to make sure that the reaction Ga-EDTA has taken place completely. This was done by heating the solution for some time at 90 °C. Check tests made with standard gallium solutions showed that equilibrium had in fact been established.

The total uncertainty in the analysis — the random and the systematic uncertainties together — can be determined by subjecting reference samples of pure gallium and of pure gallium with pure selenium to the same procedure as the samples analysed.

By combining various analytical techniques we were able to show that the total content of impurities in the reference samples was of the order of 10^{-5} wt%. We also established that the interference due to the presence of selenium caused a systematic uncertainty of less than 0.01 %, i.e. less than the random uncertainty of the analyses.

Since the same gallium and selenium were used for preparing the gallium selenide and the reference

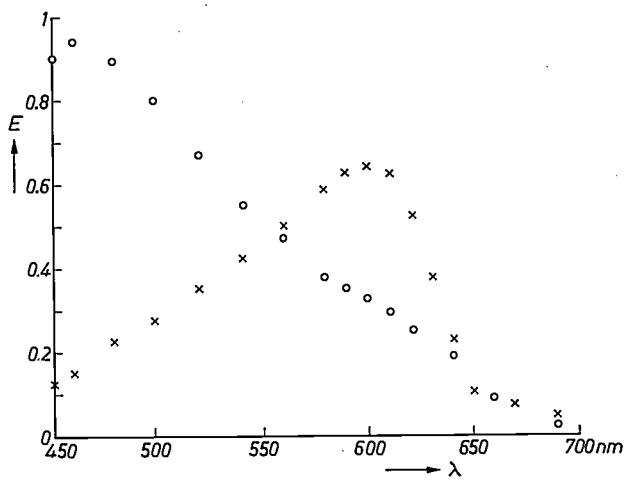


Fig. 5. Absorption spectra of the free indicator methyl thymol blue (MTB) (circles) and of the complex that gallium forms with this indicator (crosses). The complex formation is accompanied by a marked change in absorption. The absorption E, plotted on the ordinate, is defined as $\log_{10}(I_0/I)$, where I_0 is the intensity of the incident light and I the intensity of the transmitted light. The change of absorption due to the addition of ethylene diamine tetra-acetic acid (EDTA), which displaces MTB from the complex, has its maximum at a wavelength of about 600 nm, which is therefore the wavelength at which the measurements can best be carried out.

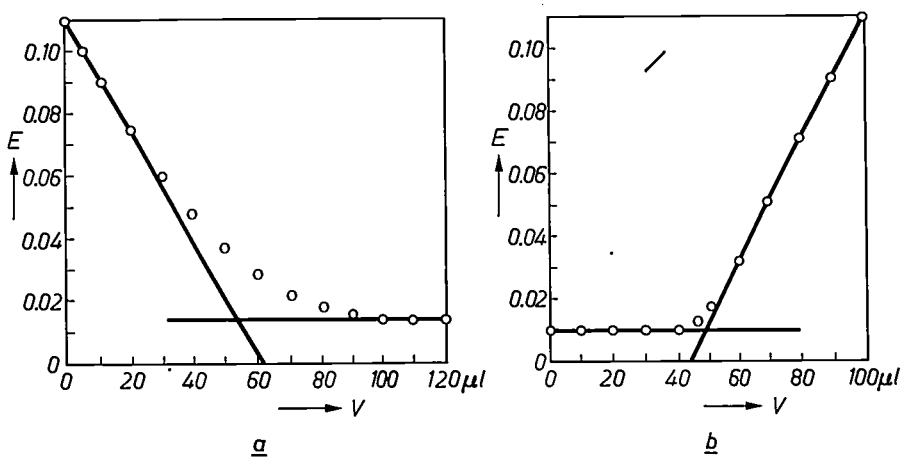


Fig. 6. The absorption E measured photometrically at a wavelength of 600 mm, as a function of the volume V of titrant added in the end-point detection phase. *a*) Direct determination of gallium by titration with EDTA. The absorption first decreases and later becomes constant. *b*) Indirect determination of gallium by titration of a small excess of EDTA with zinc. When the titrant is added the absorption initially remains constant, and then shows a marked linear increase. The end-point is taken to be the point, found by extrapolation, where the straight parts of the curve intersect.

samples, the only uncertainty remaining is the possible contamination introduced during the actual preparation. This could cause an error in the interpretation of the initial weighing and could also affect the photometric titration. Comparative analyses, however, showed that there were no impurities in higher concentrations than the order of the uncertainty in the analysis.

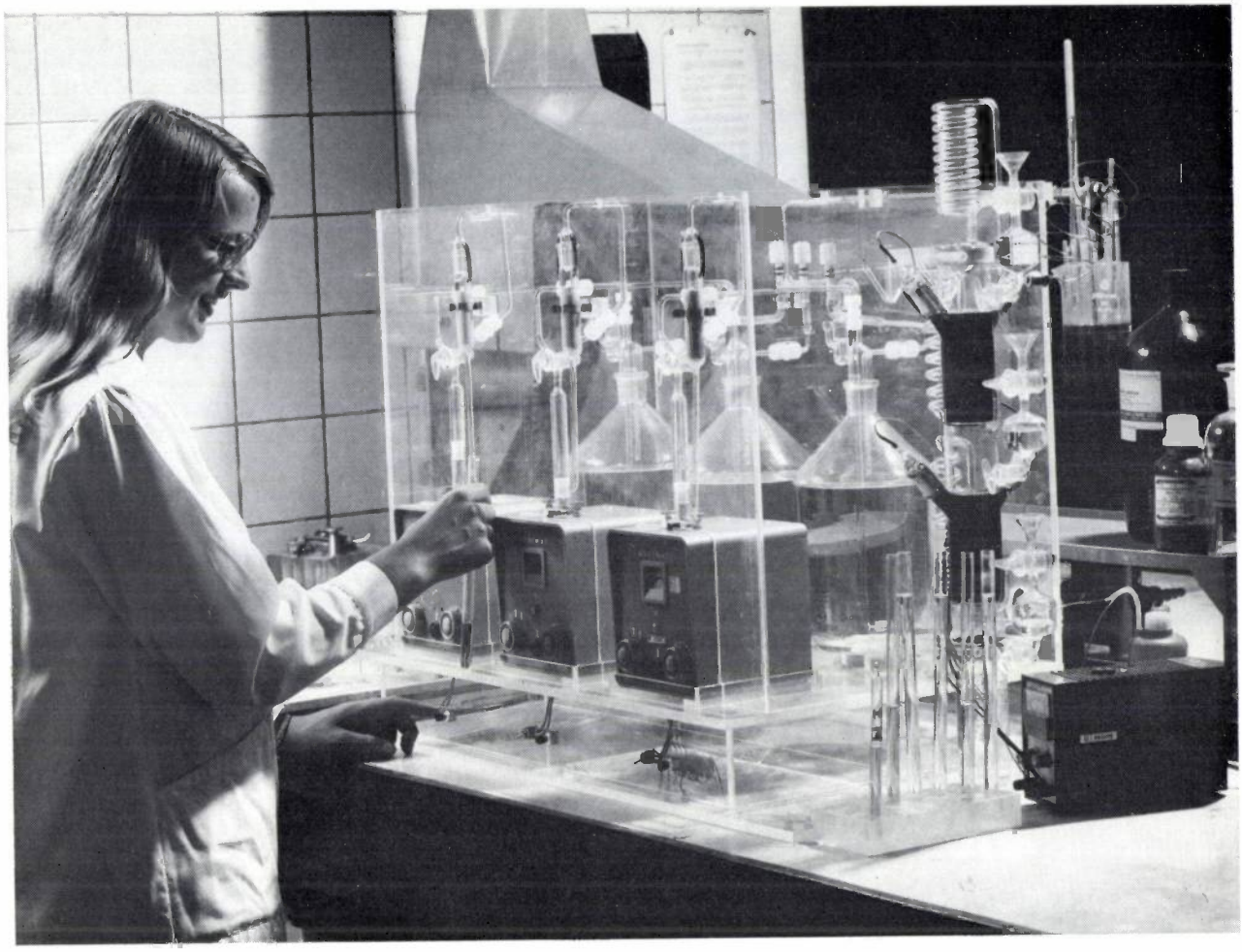
Table I summarizes the results of the three types of analyses that we carried out. As can be seen, with neutron activation analysis a total uncertainty of the order of 0.1% can be achieved. X-ray fluorescence analysis can be made almost as accurate as neutron activation analysis and as we mentioned earlier, it has the advantage that it is considerably faster. Further study of the sources of error is needed, however.

The complexometric determination of gallium in gallium selenide shows a greater standard deviation than was found in the analysis of the standard samples of gallium and of gallium and selenium. This points to some inhomogeneity in the material.

Table I. Stoichiometric analysis of gallium selenide by neutron activation analysis, X-ray fluorescence analysis and titrimetric analysis.

	Neutron activation		X-ray fluorescence		Titrimetric analysis
	Ga	Se	Ga	Se	Ga
Average content wt%	41.74	58.19	41.77	58.19	41.70
Standard deviation	0.06	0.08	0.08	0.06	0.04

Summary. A thorough investigation of the sources of error in all the steps of the analytical procedure made it possible to use neutron activation for a stoichiometric analysis of gallium selenide with a total uncertainty of 0.1%. The measurements were checked by X-ray fluorescence and also by means of a refined titrimetric determination of the gallium content. A total uncertainty of the order of 0.1% was also obtained in the X-ray fluorescence analysis. The main sources of errors in neutron activation analysis are the irradiation conditions, which are difficult to reproduce exactly for both analysis and reference samples. Calibration with reference samples whose impurity content was less than 10^{-5}% showed that the total uncertainty in the titrimetric determination was less than 0.01%. The greater standard deviation found in the determination of gallium in the selenide (0.1%) indicates some inhomogeneity in the material.



Determining the content of active oxygen in oxides

The photograph shows the accurate metering of a quantity of ferrous solution of known concentration for determining the content of active oxygen in oxide mixtures, such as ferrites. The solution also contains hydrochloric acid (6M) to dissolve the sample. Depending on the expected content of a sample, a choice is made from the three solutions (ferrous concentrations 10^{-3}, 10^{-2} and 10^{-1} M) in the three storage bottles standing at the back of the bench. The solutions are transported in electrically operated metering burettes.

To avoid oxidation, all the ferrous solutions are stored and transported in an inert-gas atmosphere. To keep the inert gas (argon or nitrogen) free from oxygen it is allowed to bubble through a solution of viologen[1], an organic substance that binds oxygen very effectively. The viologen solution, which is blue in the reduced state and colourless in the oxidized state, is contained in two small vessels on the right of the apparatus. The solution can again be reduced electrolytically. The current source used for this reduction can be seen in the foreground.

This procedure in an analysis is as follows. The sample is introduced into a tube that has previously been flushed with the inert gas, the 6M hydrochloric acid solution containing a known concentration of ferrous ions is added and the tube is sealed off. The tube is now heated for some time at a temperature between 100 and 200 °C, which dissolves the compounds to be analysed. If they contain active oxygen, Mn, Co, Ni or Pb in a trivalent or tetravalent form, part of the ferrous ions will be oxidized to ferric ions. The residue of ferrous ions is found by coulometric titration with a solution of Ce^{4+} ions. The same determination is done on a blank ferrous solution and the active oxygen content in the analysed material is found from the difference between the two determinations. If the material analysed contains divalent iron, a higher ferrous-ion content is found than in the blank determination.

[1] R. E. van de Leest, Electro-anal. Chem. and Interfacial Electrochem. 43, 251, 1973.

Surface analysis, methods of studying the outer atomic layers of solids

H. H. Brongersma, F. Meijer and H. W. Werner

Introduction

Information on the atomic composition and structure of a solid can be obtained by studying the interaction of a beam of atoms, ions, electrons or photons with the surface of the solid. The thickness of the layer for which information is obtained is determined by the depth from which emitted particles can escape from the surface and be detected. In all cases the experimental conditions are chosen to promote strong interaction between the solid and the primary or secondary particles, and therefore the investigated layer is only a few atomic layers thick. This layer, with a thickness of up to about 5 nm, is what is meant in this article when we refer to the surface of a solid.

The methods we shall discuss here are particularly sensitive for the outer atomic layers of a solid, and are therefore said to possess a high 'surface sensitivity'. This means that the measurements often have to be carried out in an ultra-high vacuum to prevent the results from being affected by adsorbed gases. At a gas pressure of 10^{-4} Pa (about 10^{-6} torr), and given a sticking probability of 1, a monoatomic adsorbed layer forms in one second. To allow reasonable time for measurements on a clean surface the pressure must therefore be at least factor of 10^3 lower. On many surfaces, and certainly those for 'technical' applications, the sticking probability is very much lower (by a thousand times or more), so that the pressure can be correspondingly higher. To obtain information at a greater depth in the sample it is usually necessary to remove thin layers of material from the surface. This is generally done by bombardment with fast ions, which gives sputtering.

To characterize a surface it is important in the first place to have qualitative and quantitative information on the composition, and it may sometimes be desirable to know the variation of the composition along the surface. Other parameters are the structure of the surface, i.e. the relative geometrical arrangement of the surface atoms, the chemical bonding between these atoms and with the atoms in the bulk, and the location of the energy levels for the electrons in the surface layer. At the surface these structures usually differ from those in the bulk [1]. Finally there is the macroscopic structure, the surface roughness.

Dr H. H. Brongersma, Dr F. Meijer and Dr H. W. Werner are with Philips Research Laboratories, Eindhoven.

Table I. The various possibilities for surface analysis through the interaction between the outer atomic layers and a beam of ions, electrons or photons. In the investigation the ions, electrons and photons emitted in the interaction are detected. The techniques discussed in the article are shown in black, others in grey.

particles in / out	ions	electrons	photons
ions	SIMS NIRMS	ESD	
electrons	INS	LEED AES ELS	ESCA(XPS) UPS
photons	IIL	APS	ellipsometry

SIMS	Secondary ion mass spectrometry
NIRMS	Noble-gas-ion reflection mass spectrometry
INS	Ion-neutralization spectroscopy
IIL	Ion-induced light emission
ESD	Electron-stimulated desorption
LEED	Low-energy electron diffraction
AES	Auger electron spectroscopy
ELS	Electron energy-loss spectroscopy
APS	Appearance-potential spectroscopy
ESCA	Electron spectroscopy for chemical analysis
XPS	X-ray photo-electron spectrocopy
UPS	Ultraviolet photo-electron spectroscopy

All this information is relevant to the study of processes that take place at the surface of a solid. A typical example is catalysis, which involves very specific reactions with surface atoms or groups of atoms. Corrosion, which eats up so much money, is another process that can be better understood by means of surface investigations, with the hope of finding more efficient methods of combating it. Knowledge of surfaces is also important in many areas of technology, as in planar semiconductor techniques and for the investigation of cathodes in thermionic tubes.

The principal methods of surface analysis by means of bombardment with ions, electrons or photons are summarized in *Table I*. Other methods used for the investigation of surfaces, such as field emission of electrons and ions [2], are less suitable for general application in surface analysis and will not therefore be considered here. The first section reviews the analytical methods in which a sample is bombarded with electrons and photons. The next section deals with the

[1] M. J. Sparnaay, The electrical double layer, Pergamon Press, New York 1972.
[2] A. van Oostrom, Philips tech. Rev. 33, 285, 1973.

analytical methods using ion bombardment. The article concludes with a short comparative survey of the analytical methods described and of the information they are capable of providing.

Surface analysis using electrons and photons

We shall discuss here some surface-analysis techniques in which the sample is bombarded with electrons or photons, giving rise in turn to the emission of electrons and photons. These are the methods mentioned in the four compartments at the lower right in Table I. Although the interactions of electrons and photons with a solid are in some ways very similar, an important difference is that the total cross-section for inelastic scattering is very much greater for low-energy electrons (< 10 keV) than for photons of comparable energy. If electrons originating from a particular scattering process, and possessing an energy characteristic of that process, are to be detectable as such, they must lose no more energy in the solid on their way to the energy analyser. This implies that the surface sensitivity of the methods discussed here is due to the short mean free path of the electrons involved in the process (see *fig. 1*). An exception to this, as we shall see, is ellipsometry.

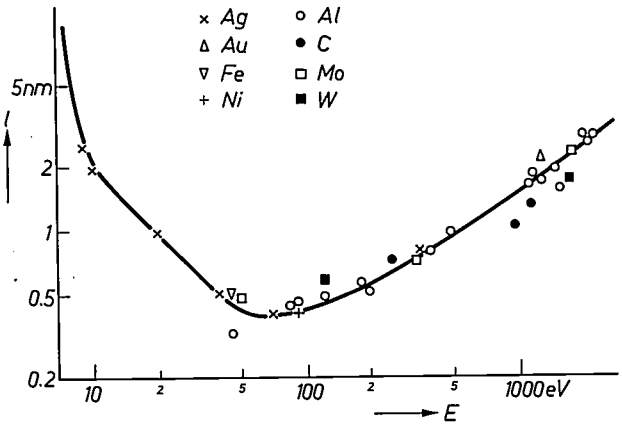


Fig. 1. The mean free path l of an electron in various solids as a function of the energy E of the electron. The main process by which electrons lose energy in a metal is the excitation of plasma oscillations.

Elastic scattering of low-energy electrons

Electrons are scattered in a solid both elastically and inelastically. A method of surface analysis based on elastic scattering is low-energy electron diffraction (LEED). A sample is bombarded with electrons of energy varying from 10 to 500 eV and the emitted electrons are selected by an energy analyser arranged so that only those that are elastically scattered are detected. The apparatus most commonly used for this purpose is illustrated schematically in *fig. 2*. The wavelength λ of a particle with velocity v and mass m (the De Broglie wavelength) is given by the equation:

$$\lambda = \frac{h}{mv}. \qquad (1)$$

The wavelength in nm of electrons that have travelled through a potential difference V is given by $\lambda = \sqrt{(1.5/V)}$, where V is expressed in volts. The wavelength of electrons with energy of the order of 100 eV is thus of the same order of magnitude as the atomic distances in the solid. If the sample is a single crystal, strong diffraction effects will appear which, because of the considerable inelastic scattering (see fig. 1), remain limited to the outer atomic layers. As remarked earlier, the scattering cross-sections for photons are much smaller, and therefore X-ray diffraction (photons with the same wavelength as the electrons treated here) is a

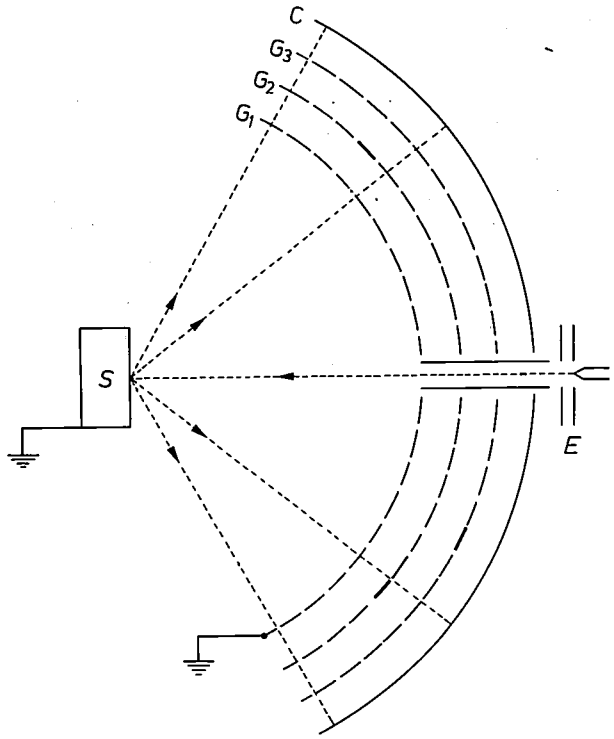


Fig. 2. Arrangement for low-energy electron diffraction (LEED) and Auger electron spectroscopy. The sample S is bombarded with electrons from the electron gun E. Around the sample are situated four concentric spherical electrodes whose centres of curvature coincide with the point where the electron beam strikes the sample. Three of these electrodes are the grids G_1, G_2 and G_3; the fourth electrode C serves in LEED as the fluorescent screen and in Auger spectroscopy as the collector. G_1 and S are always at earth potential, so that the space between them is field-free; the electrons thus travel in straight lines in the escape direction after leaving the sample. The two other grids are employed for energy selection. In LEED the electrons are subsequently post-accelerated to ensure that they strike the fluorescent screen with an energy high enough to produce an adequate light output. In Auger spectroscopy these grids generate a variable opposing field, used for analysing the energy distribution of the Auger electrons.

method for bulk studies. Diffraction effects give information about structure. The information that can be obtained from LEED, like that from X-ray diffraction, can be divided into the dimensions of the unit cell, which is derived from the direction of the diffracted beams, and the internal structure of this cell, which must be calculated from the relative intensities of the diffracted beams. The first information is easily obtained, but determining the positions of the atoms in the unit cell is not nearly as simple. What is more, the theoretical problems of determining these positions are much greater for LEED than for X-ray diffraction, because of the very strong interaction of the electrons with the solid [3]. Recently, however, considerable progress has been made in this field, and some structures at the surface have been solved, the positions of the atoms having been determined with reported accuracies of 0.01 nm. This can only be done, however, for very simple structures, whereas with X-ray analysis exceptionally complex crystal structures can be determined.

Most LEED experiments only give the dimensions of the surface unit cell, and this information has given rise to some speculation about possible surface structures. As long ago as 1959 it was found that the unit cell on a clean silicon surface was not identical with the unit cell of an ideal lattice. This implies that there has been a rearrangement of the surface atoms to produce a more stable configuration. Since then many such deviating surface structures have been found, especially in semiconductors. Often, the adsorption of foreign atoms to clean surfaces leads to new unit cells on the surface. This is illustrated by the LEED recordings in *fig. 3*, showing a germanium surface before and after the adsorption of sulphur [4].

Starting from the dimensions of the unit cell at the surface, and taking chemical data such as bond lengths, attempts can be made to build up a model of the surface structure. *Fig. 4* gives as an example a possible structure for sulphur adsorbed on a germanium surface, based on the LEED recordings shown in fig. 3. The LEED method is also sensitive to surface roughness, permitting in particular the accurate determination of ordered steps of atomic dimensions, and also the average displacements of the surface atoms caused by thermal motion of the lattice.

Auger electron spectroscopy

In the middle of the 1960s research on surfaces received fresh impetus when it was found that a slightly modified LEED apparatus could provide a quantitative atomic analysis of the surface composition in a relatively simple way. The new method was based on a process discovered in 1925 by P. Auger in research on gases, and has been called Auger electron spectroscopy (AES).

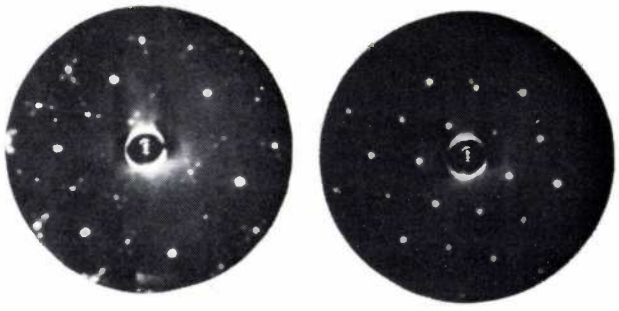


Fig. 3. LEED recordings of the (111) face of germanium, made with an electron energy of 60 eV. *Left*: the diffraction pattern of the clean surface. *Right*: the pattern after adsorption of sulphur. The appearance of spots half-way between the original ones indicates the formation of a structure on the surface with a period equal to twice the period of the (111) face in a germanium crystal.

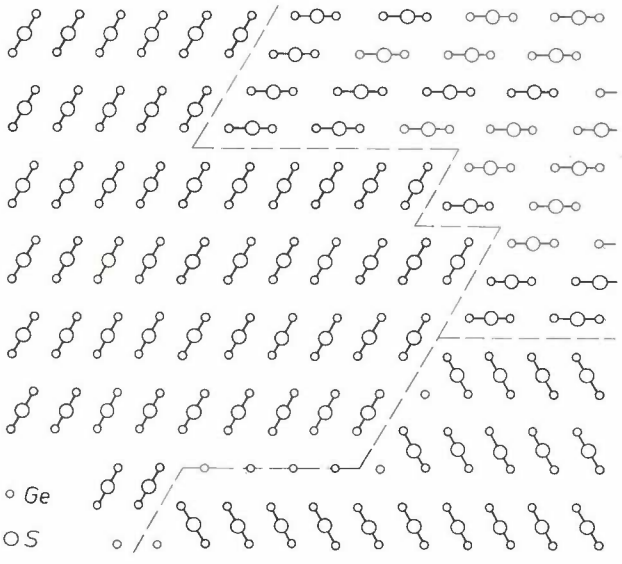


Fig. 4. Structure of the (111) face of germanium covered with a monoatomic layer of sulphur. Each germanium atom at the surface has a 'dangling' bond; two of these bonds may be saturated by one sulphur atom. A sulphur structure is thus formed whose period is twice that of the surface structure of germanium. Since the sulphur adsorption begins simultaneously at different places, domains form on the surface that do not fit together. The boundary between two such domains is indicated by a dashed line. The occurrence of three different orientations for this divalent surface structure gives rise to the observed trivalent symmetry.

In LEED only the elastically scattered electrons are detected, whereas in Auger spectroscopy the whole spectrum of inelastically scattered electrons is recorded as well, as shown in *fig. 5a*. In addition to the elastically scattered electrons and a number of electrons that have undergone characteristic energy losses up to about 30 eV (called plasma and interband losses) there is a broad, high background of electrons that have lost energy in various ways, usually as a result of a succession of processes. Superimposed on this background

[3] C. B. Duke, N. O. Lipari, G. E. Laramore and J. B. Theeten, Solid State Comm. **13**, 579, 1973.

[4] A. J. van Bommel and F. Meijer, Surface Sci. **6**, 391, 1967.

are a number of small peaks, called Auger peaks. Since these peaks are small variations on a relatively high slowly varying background, it is difficult to enhance them by direct amplification. The Auger peaks are usually separated from the background by electronic differentiation of the energy-distribution curve. A typical result can be seen in fig. 5*b*, which shows, as we shall see presently, that each Auger peak on the energy scale corresponds to a particular kind of atom.

place in a process known as an Auger transition. In this transition the hole in the inner shell is filled by an electron from a higher band, and the energy released is transferred to another electron in this band. This electron now has sufficient energy to be emitted and leaves the solid with an energy characteristic of the energy-level diagram of the atom, and hence of the type of atom. The position of the Auger peak is largely determined by the energy of the inner shell, and nearly all

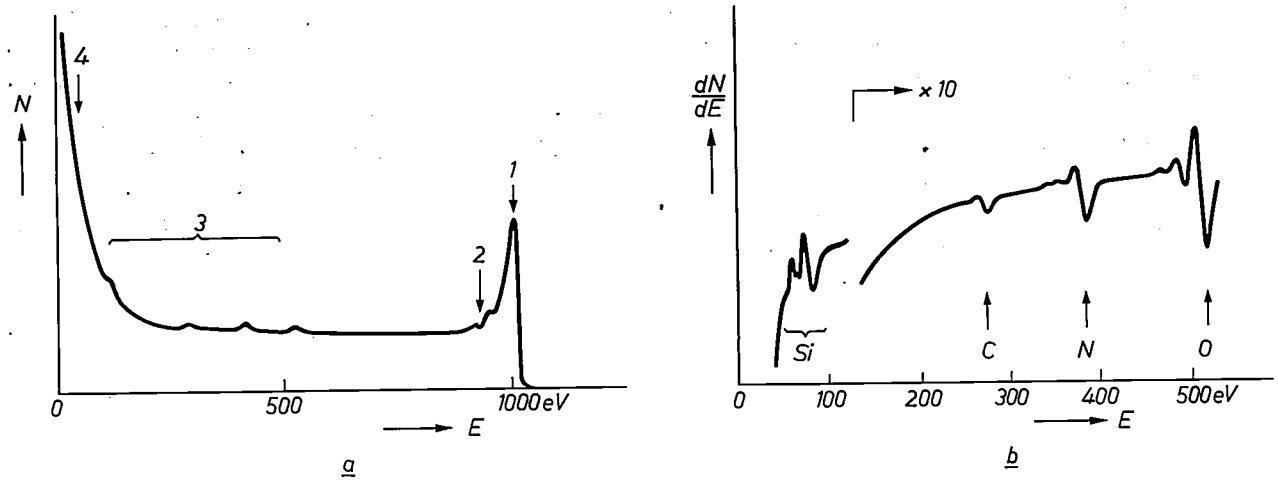


Fig. 5. *a*) Energy spectrum of the back-scattered electrons in electron bombardment of a solid. The number of electrons N of energy E is plotted as a function of the energy. *1* elastically scattered electrons. *2* structure arising from energy losses due to the excitation of plasma oscillations and inter-band transitions. *3* Auger peaks on a relatively high background. *4* true secondary electrons. *b*) Auger spectrum obtained by single differentiation of the energy spectrum of the back-scattered electrons. The spectrum relates to a silicon surface layer consisting of silicon nitride and silicon oxide. Half of a monolayer of carbon is observed on the surface of this layer. The elements corresponding to the peaks in the spectrum are marked below them. The thickness of the oxide-nitride layer was determined by ellipsometry and was found to be 4.4 nm.

It can very occasionally happen that Auger peaks corresponding to different elements overlap.

The process giving rise to the Auger electrons is illustrated schematically in *fig. 6*. When an electron enters the surface layer there is a certain probability, characterized by the ionization cross-section σ, that it will ionize an atom in the sample, thereby causing the ejection of an electron from an inner shell of the atom. After some time, which is long compared with the time needed for the ionization, de-excitation of the atom occurs and an electron falls from a higher electron level of the solid into the inner shell. Because of this relatively long time interval the de-excitation is not affected by the manner in which the ionization was brought about. The energy released during the de-excitation may be emitted as an X-ray quantum. If, however, the energy of the inner shell is less than a few keV below the vacuum level, X-ray emission is very unlikely. This is always the case in Auger electron spectroscopy, which is usually performed with primary electrons of energy of no more than 5 keV, and de-excitation then takes

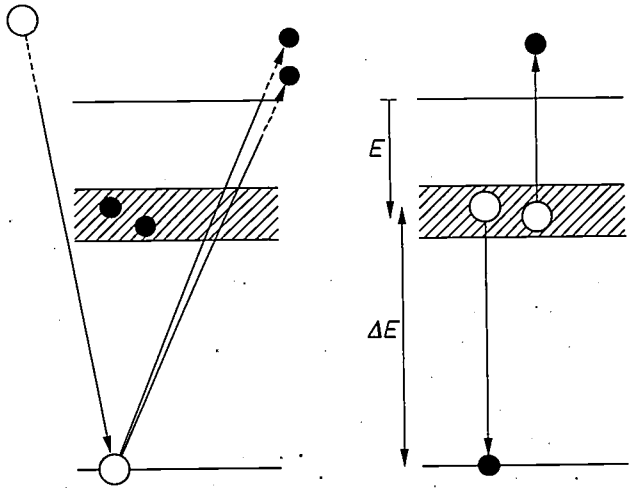


Fig. 6. Schematic illustration of the Auger transition, based on the energy-level diagram of the electrons in a solid. *Left*: a hole forms in an inner shell of one of the surface atoms as a result of an incoming electron. *Right*: the hole is filled by an electron from a higher level; the energy thus released is transferred to another electron, which is then emitted. The energy of the emitted electron is equal to the difference of the energies E and ΔE and is thus characteristic of the energy-level diagram of the element in question.

peaks may be unambiguously interpreted in terms of kinds of atom. The chemical environment of the particular atom in the sample has a relatively small effect on the position of the energy levels and hence on the energy of the Auger electrons (called the chemical shift). As we noted earlier, the energy of the primary electrons has no effect at all on the energy of the Auger electrons. The surface sensitivity of the method is attributable to the short mean free path mentioned earlier of low-energy electrons. The Auger electrons in particular determine the depth of information, for if they have lost energy in the solid they are no longer detectable as such, but merge into the general background.

There are two methods for the quantitative interpretation of Auger spectra. The first is based on a comparison of Auger peaks (usually measured as the peak-to-peak height in the $dN(E)/dE$ spectra, fig. 5*b*) with calibration spectra measured on pure elements or compounds. This often yields a reasonable value for the concentration of a particular kind of atom at the surface of a sample, and in less favourable cases it does at least give the order of magnitude. This method is therefore the one most commonly used. The other method consists in measuring the Auger electron current i_A at the collector of the energy analyser and interpreting it in terms of the parameters that describe the Auger transition [5]. Since this description gives some understanding of the physical background of the Auger transition, we shall consider it in somewhat more detail.

For the simplest case of a sample which has N foreign atoms per cm^2 in its outer atomic layer, the measured Auger electron current is given by:

$$i_A = \sigma i_p N \operatorname{cosec} \phi_p \cdot Tqr, \qquad (2)$$

where i_p is the total primary electron current at the sample and σ is the ionization cross-section. The term $\operatorname{cosec} \phi_p$, where ϕ_p is the angle of incidence of the primary electron beam on the sample, appears in the equation because the primary electron beam at glancing incidence 'sees' a relatively greater density of surface atoms. T is the transmission factor of the energy analyser and indicates the ratio of the electrons collected to those emitted from a sample that is assumed to be an isotropic emitter. The quantity q denotes the escape probability of the Auger electrons. In the case considered here it may be assumed that an Auger electron emitted in the direction of the analyser is in fact able to escape from the upper atomic layer ($q = 1$). For an atom in a layer below the surface this is not the case; q then depends directly on the mean free path for inelastic scattering in the sample. Finally, r is called the back-scattering factor. This allows for ionizations not due to the primary electrons but to electrons that are back-scattered inside the solid and pass through the surface layer again on their way back. This contribution to the Auger electron current will often be between 10 and 60%, corresponding to r values of 1.1-1.6. Published values are to be found for the parameters σ, q and r, while i_p, ϕ_p and T can be determined experimentally. Thus, the measurement of i_A gives a direct value for N, the density of foreign atoms at the surface of a sample.

The situation becomes rather more complicated when the foreign atoms are present not only in the upper layer but are distributed over the depth from which the Auger process can supply information. There are various methods, however, by which some idea can be obtained about the distribution in the deeper layers. The different Auger peaks of an element can be measured, and since the electrons that cause these peaks have different energies they also have different mean free paths in the solid. Another method is to vary the energy of the primary electrons, thus varying their effective depth of penetration. In both cases the effect is to vary the Auger signal from atoms in deeper layers with respect to that from atoms in the upper layer. The rest of the information is interpreted in the same way as described with equation (2).

Nowadays there are also forms of Auger apparatus in which the surface of the sample is scanned with an electron beam with a typical diameter of 5 μm. In this way information is obtained on the local distribution of different kinds of atom at the surface.

Although the Auger method is essentially non-destructive, the possible effect of the electron beam on the surface should nevertheless be borne in mind [6]. The electrons may induce both adsorption and desorption of molecules. The primary electron current should not therefore be too high.

ESCA

A method closely related to Auger spectroscopy is XPS (X-ray Photoelectron Spectroscopy), more generally known as ESCA, which stands for Electron Spectroscopy for Chemical Analysis [7]. In this method the inner shell is ionized by an X-ray photon of appropriate energy. The peaks in the ESCA spectrum can have two different causes. The peaks may be due to photons transferring all their energy to the inner-shell electron, which is then emitted with a specific energy $h\nu - E_b$, where E_b is the binding energy of the electron. In addition to these peaks, there are the Auger peaks resulting from the completion of the inner shell. In Auger spectroscopy the electrons directly emitted

[5] F. Meijer and J. J. Vrakking, Surface Sci. 33, 271, 1972.
[6] B. A. Joyce and J. H. Neave, Surface Sci. 34, 401, 1973.
[7] K. Siegbahn et al., ESCA — Atomic, molecular and solid state structure studied by means of electron spectroscopy, publ. Almqvist, Uppsala 1967.

from the inner shell do not give sharp peaks. This is because the available energy upon ionization by electrons is 'randomly' distributed between the primary and the secondary electron.

ESCA supplies essentially the same kind of information as Auger spectrometry, that is to say it gives an atomic analysis and, provided the results of the measurement are properly interpreted, a reasonably good quantitative analysis of a surface layer whose thickness is again determined by the escape depth of the emitted electrons in the solid. As in Auger spectroscopy, the position of the peaks is affected by the chemical environment of the atoms. The chemical shift that this causes in ESCA is much easier to interpret, since there is only one energy level involved in the process. With ESCA it is therefore possible to discern chemical differences, such as that between sulphur in a sulphate and in a sulphide.

Ultraviolet photoelectron spectroscopy and ellipsometry

The precise energy at which the Auger or ESCA electrons are detected, and the shape of the peaks in the spectrum, provide information on the chemical environment of the atom from which the emission takes place or, in other words, on the local states of the valence electrons. Information on the electron states in the surface layer, a kind of surface band structure, can be obtained by ultraviolet photoelectron spectroscopy (UPS) [8]. Electrons from these surface states are excited by photons with an energy of 5-50 eV to energy levels above the vacuum level, at which they can be emitted. The only difference compared with ESCA really lies in the energy of the photons used. The energy of the emitted electrons, less the photon energy, gives the energy position of the levels from which they originate. States in the bulk of the material as well as at the surface are measured in this way. At higher photon energies the energy of the excited electrons from a particular level will also be higher, and because of the shorter mean free path at this higher energy (see fig. 1, to the left of the minimum) the surface effects will be relatively enhanced.

Finally, a few words about ellipsometry, a method in which the surface sensitivity does not depend on the short mean free path of the particles [9]. This is an optical reflection technique that makes use of the change in the state of polarization of polarized monochromatic light on reflection at a surface. With optical components of glass or silica the wavelength of the light can be varied between 250 and 2500 nm (photon energies between 5 and 0.5 eV). The reflection of light takes place at an optical discontinuity, i.e. an abrupt or gradual change in the dielectric constant. An optical discontinuity is always to be found at a surface, e.g. the vacuum-solid interface. If the surface of the sample is covered with a layer of, say, adsorbed molecules, there are two interfaces, as shown schematically in *fig. 7*. If the path difference between the light reflected from the two interfaces lies within the coherence length, the light will contain information about the surface layer. The information consists of the thickness of the layer and its optical properties (absorption coefficient and refractive index). These optical properties can be translated into an electron-energy-level diagram. Ellipsometry is normally used, however, to measure layer thicknesses between 0.01 nm and 0.5 μm, an application that does not come within the scope of this article.

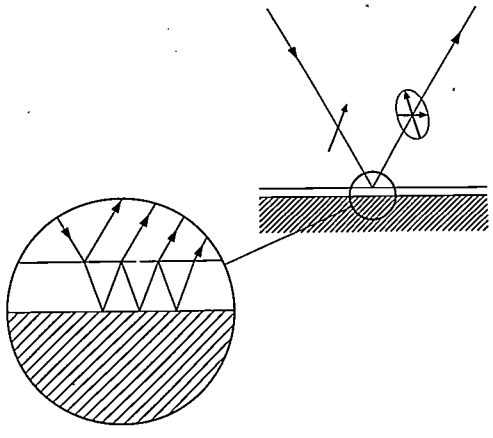


Fig. 7. Schematic representation of ellipsometry. The incident light is linearly polarized. On reflection from the boundary faces of a surface layer its state of polarization changes. The combined action of all the reflections within the coherence length of the incident light gives the emergent beam an elliptical polarization. The degree of ellipticity and the orientation of the ellipse are determined by the refractive index, the absorption coefficient and the thickness of the surface layer.

Surface investigation with ions

Ion scattering at surfaces

The interaction of an ion beam with a solid surface may involve several processes (*fig. 8*). One of these is scattering of the primary ions, which is essentially a complicated process. It can be greatly simplified, however, by careful choice of the experimental conditions. There are many features of ion scattering that can be explained quite satisfactorily by treating the scattering in terms of classical mechanics as an elastic collision [10]. This is because the De Broglie wavelength λ of ions with energies greater than the thermal energy is small compared with the atomic distances in the surface. For Ne^+ ions with an energy of 1 keV it follows from equation (1) that $\lambda = 0.0002$ nm, whereas the lattice spacings in the solid are much greater, generally amounting to some tenths of an nm. The ions thus only 'see' isolated atoms. There is a certain probability that ions striking the surface will leave the crystal again after one or more collisions. It is possible to study the scattering of ions

from a single surface atom by using inert-gas ions with an energy of about 1 keV as primary ions.

Inert-gas ions have a very strong electron affinity. The probability that these ions will be neutralized when they collide with the surface is therefore very great. Thus, only about 0.1% of the He^+ atoms striking a surface retain their charge after colliding with the first atom. The number of ions still ionized after two or more collisions is therefore negligibly small. For other ions, however, and especially at small scattering angles, these values may be appreciably higher. Consequently, when an analyser is used that only detects ions, it will detect virtually only those particles that have had only one collision with an atom. A great advantage of this method is that at energies of about 1 keV the collision time is much shorter (10^{-15} s) than the characteristic vibration time (10^{-13} s) of a surface atom. The ion is thus already far away from the surface again before the surface atom 'realizes' that it is attached to the lattice. A good approximation to the collision of the He^+ ion with a solid surface is therefore the collision of an He^+ ion with a single free atom, a process that is easily described in terms of classsical mechanics. By applying the laws of the conservation of energy and of momentum, it is possible to find the relation between the energy of the ion before and after scattering, the scattering angle and the masses m_{ion} and m_{at} of the ion and of the atom. Taking the ions whose scattering angle is 90°, we can write this relation very simply as:

$$E_f = \frac{m_{at} - m_{ion}}{m_{at} + m_{ion}} E_i, \qquad (3)$$

provided we assume that the scattered ion is lighter than the atom ($m_{ion} < m_{at}$). Here E_f is the energy of the scattered ion and E_i the energy of the incident ion. If the ions in the primary beam are selected by mass and energy, we find that the energy distribution of the scattered ions gives information about the mass of the atoms they collided with at the solid surface. *Fig. 9* gives as an example a typical energy distribution obtained when a beam of Ne^+ ions at an energy of 300 eV was directed at a nickel crystal to which Cl, Br and I were adsorbed. The figure shows the energy distribution of the ions scattered at an angle of 90°. It is evident that this energy spectrum is in fact a mass spectrum of the atoms on the surface. The mass resolution is best when the mass of the surface atom is about the same as that of the ion. With Ne^+ ions we can thus separate the isotopes of chlorine but not those of nickel. To separate Ni into its isotopes we would have had to use Ar^+ ions. The validity of equation (3) appears from the good agreement between the observations presented in fig. 9 and the positions of these peaks calculated with the equation for this particular case. The equation states

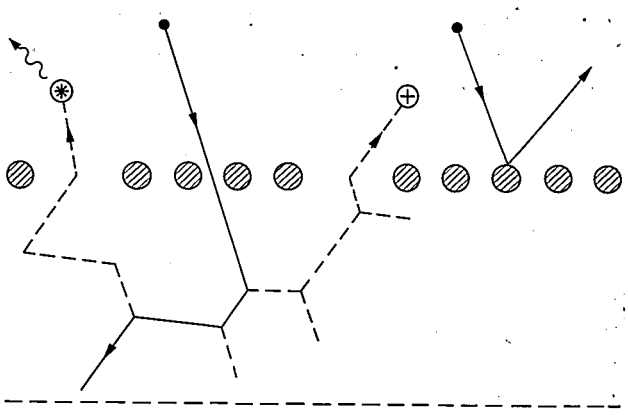


Fig. 8. Illustrating the processes that occur when a solid is bombarded with ions. *Left*: multiple collisions of the incoming ion with the atoms of the solid result in the emission of ions or excited surface atoms. De-excitation of these surface atoms leads to the emission of light or of ions. These are the processes underlying secondary-ion mass spectrometry (SIMS) and ion-induced light emission (IIL). *Right*: elastic collision of an incoming ion with a surface atom. The energy of the scattered ion is partly determined by the mass of the surface atom that takes up part of the primary energy upon the collision. This is the process studied with inert-gas-ion reflection mass spectrometry (NIRMS).

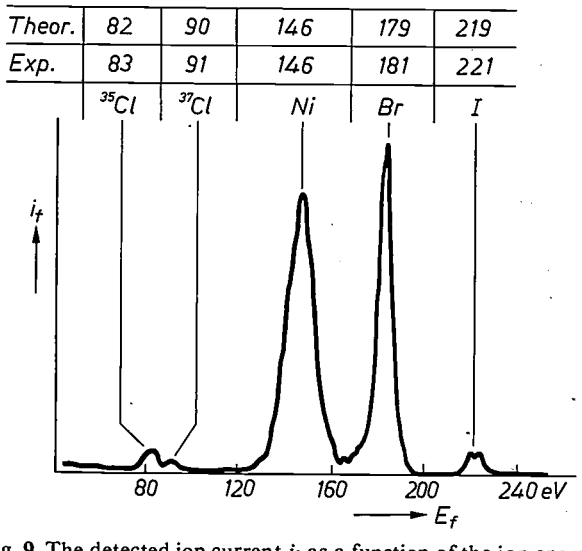


Fig. 9. The detected ion current i_f as a function of the ion energy E_f after scattering of Ne^+ ions with an energy of 300 eV from a nickel surface to which chlorine, bromine and iodine are adsorbed. The measured values of E_f and those calculated from equation (3) for the various peaks are shown above the graph. The scattering angle here was 90°.

that, given a fixed scattering angle and fixed collision partners, the ratio E_f/E_i is also constant. In *fig. 10*, where E_f is plotted as a function of E_i for the scattering of Ne^+ ions from nickel and a scattering angle of 45°, it can be seen that this is in fact the case.

Ion scattering can be measured with the instrument illustrated in *fig. 11*, which was built at Philips

[8] L. F. Wagner and W. E. Spicer, Phys. Rev. B 9, 1512, 1974.
[9] F. Meijer, Ned. T. Vacuümtechniek 11, 77, 1973. See also K. H. Beckmann, Philips tech. Rev. 29, 129, 1968, and F. Meijer and G. A. Bootsma, Philips tech. Rev. 32, 131, 1971.
[10] H. H. Brongersma, J. Vac. Sci. Technol. 11, 231, 1974.

Research Laboratories in cooperation with the Philips Industrial Equipment Division. This spectrometer, called 'NODUS' [11], has a high sensitivity because the entire cone of ions scattered over a particular angle contributes to the detector signal.

Since ion-scattering experiments are always concerned with studying the interaction of one ion with one atom, the crystalline nature of the surface is of no significance. In the case of very rough surfaces the total intensity of the scattered ions is smaller, because many ions are then unable to reach the detector. This does not however affect the relative amplitudes of the peaks.

In surface investigations it is very important to know the greatest depth from which an atom can contribute to the spectrum. To find this out we performed experiments on the (111) face of silicon both with and without a layer of bromine chemically adsorbed on it [10]. This silicon face is highly reactive and can bind one bromine atom per surface atom. The bromine atoms also have such a large diameter that they are capable of forming a closed monolayer. Since the bromine atoms are monovalent, the reaction ceases when the monolayer is formed. Ion scattering reveals (*fig. 12*) that when the reaction has ceased the bromine completely covers the silicon. The method can thus be used selectively for the outer atomic layer.

Interpreting the spectra is relatively simple since no weighting factors need to be assigned to the deeper layers. *Fig. 13* shows another example of the surface sensitivity of the method. We analysed a nickel crystal before and after adsorption of sulphur. The system Ni-S is of great practical importance, because nickel is a catalyst used in many industrial processes and because sulphur poisons the catalytic activity of the nickel. Various models have been proposed to explain the structure of the adsorption complex. Since hardly any nickel is observable after the adsorption of sulphur, a model containing both sulphur and nickel in the upper layer, referred to as a reconstructed model, must be rejected. Combination of the measurement described above with the results of work done by M. Perdereau and J. Oudar [12] proves that the crystal surface is covered exactly by a monoatomic layer of sulphur.

The examples so far discussed have only been concerned with single scattering. In general this is found for He^+ ions, and for heavier inert-gas ions only at large scattering angles. Multiple scattering is increasingly found at smaller scattering angles or when heavier inert-gas ions are used. Low-energy ions may not even be able to collide with one atom without at the same time coming into contact with neighbouring atoms. *Fig. 14* shows an example involving both single and double collisions. The experiment related to the {111} faces of a zinc-sulphide crystal bombarded with

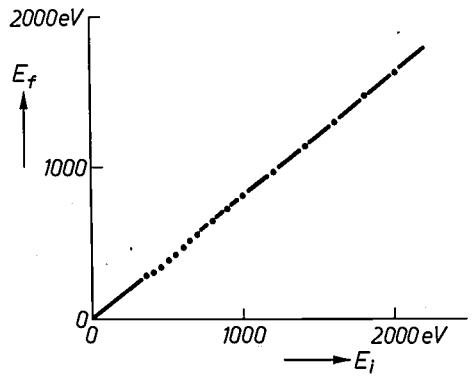


Fig. 10. Measured values of the energy E_f of Ne^+ ions scattered from a nickel surface as a function of the primary energy E_i. The ions are scattered over an angle of 45°. As expected from of equation (3), the experimental points lie on a straight line. The slope of this line is 0.81; the equation predicts that the slope would be 0.814.

1000 eV Ne^+ ions. The scattering angle was 45°. Theoretically the upper layer of the (111) face should consist only of zinc atoms, and the upper layer of the opposite $(\bar{1}\bar{1}\bar{1})$ face should consist entirely of sulphur atoms. The scattering experiments show clearly that this is in fact the case. On the one surface only single and double collisions with sulphur atoms take place, while at the other surface only single and double collisions with zinc atoms are found. Ion-scattering experiments thus enable the polarity of these crystals to be determined. This has not yet been possible with other methods of surface analysis. After etching the two faces

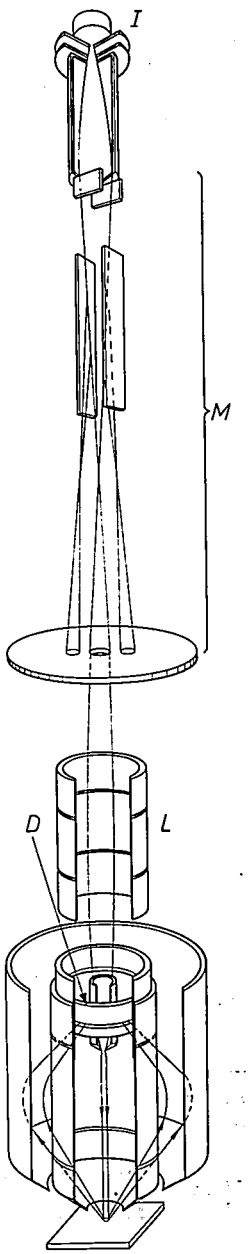


Fig. 11. Diagram of the non-destructive and ultra-sensitive single atomic layer surface spectrometer 'NODUS'. *I* ion source. *M* mass-selection system with crossed electrical and magnetic fields. The beam of mono-energetic ions leaving this system is concentrated by electrostatic lenses *L* on to the surface of the sample. All ions scattered over a particular angle with respect to the incoming beam are concentrated after energy selection on the ring detector *D*.

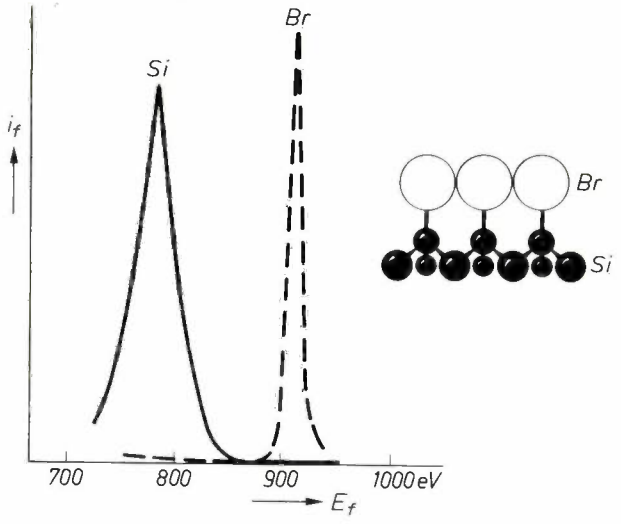


Fig. 12. Detected ion current i_f as a function of the energy E_f for Ne^+ ions scattered from a silicon surface, before and after adsorption of a monolayer of bromine. For every Si atom one Br atom is adsorbed, and the diameter of the Br atoms is such that they only just form a close-packed layer. The absence of any Si peak in the spectrum after Br adsorption proves that ion scattering in this case provides information relating only to the upper atomic layer.

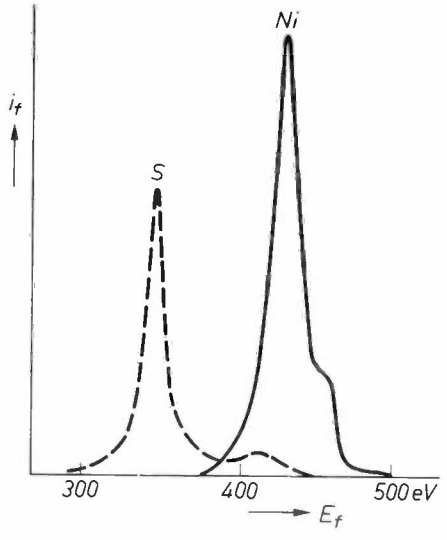


Fig. 13. Detected current i_f as a function of the ion energy E_f for ions scattered from a nickel crystal doped with sulphur. The solid curve relates to a surface cleaned by sputtering, the dashed curve relates to the same curve after heat treatment at 825 °C. The surface is now covered with sulphur, which arrives at the surface by diffusion from inside the crystal.

the chemist does see a considerable difference between them (*fig. 15*), but he cannot tell from this which surface is which. It used to be the practice to determine the absolute configurations of such non-centrosymmetric crystals and of asymmetric molecules by an X-ray analytical method known as anomalous X-ray dif-

[11] 'NODUS': non-destructive and ultra-sensitive single atomic layer surface spectrometer.
[12] M. Perdereau and J. Oudar, Surface Sci. **20**, 80, 1970.
[13] H. H. Brongersma and P. M. Mul, Chem. Phys. Letters **19**, 217, 1973.

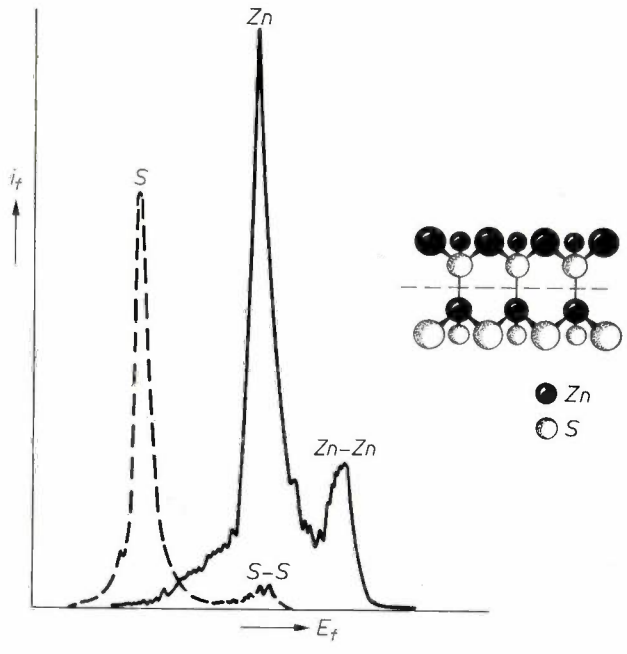


Fig. 14. *Left*: detected current i_f as a function of the ion energy E_f for Ne^+ ions scattered from two corresponding {111} faces of a ZnS crystal. The (111) face consists entirely of zinc atoms, the $(\bar{1}\bar{1}\bar{1})$ face consists exclusively of sulphur atoms. The spectra show that both single and double collisions have taken place here. *Right*: schematic structure of ZnS explaining the formation of surfaces consisting only of one kind of atom. The dashed line indicates the (111) face.

fraction. When J. Tanaka asserted in 1972 that an error in the sign had been made in all structures previously determined by this method, our study of ion scattering in ZnS could be used to show on experimental grounds that his assertion was incorrect [13].

The double collision peaks provide information about the structure of the surface. To determine the exact position of the surface atoms with the aid of these peaks it is necessary to assume a model for the interaction between ions and atoms, since the strength of the interaction potential affects the path of the ion and hence the position at which the next atom will be encountered. It was not necessary to do this for single collisions, because the interaction potential does not affect the energy of the scattered ions. By applying equation (3) successively to two collisions and using a

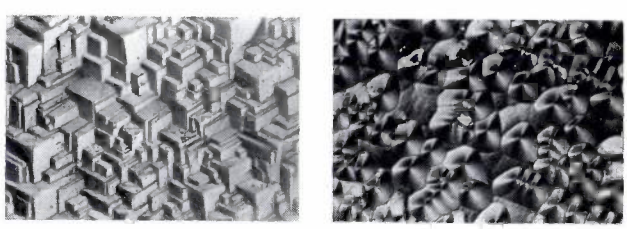


Fig. 15. Etched (111) and $(\bar{1}\bar{1}\bar{1})$ faces of ZnS. They are clearly quite different, but with etching methods it is not possible to determine whether the outer surface layers consist of Zn or S atoms.

likely model for the potential, it can be shown that the zinc and sulphur atoms in the surface of a ZnS crystal possess, within the accuracy of our measurement, exactly the ordering that would be expected from the three-dimensional structure. Multiple-ion scattering experiments also permit information to be obtained about the geometric details of a surface, such as lattice defects [14].

We can therefore conclude that, provided the experimental conditions are properly chosen, ion scattering can be used for both mass and structural analysis of surfaces. Unfortunately little can be said about the absolute quantitative accuracy to be achieved with this method. The principal uncertain factor is the neutralization probability. The information available indicates that the chemical environment of the elements to be analysed does not have much influence. The variation in sensitivity from one element to another is also found to be less than a factor of ten.

Stimulated light emission

The bombardment of a solid surface with fast ions or neutral particles gives rise to a number of emission processes in addition to the scattering phenomena already discussed (fig. 8).

Fragments leaving the surface in an excited state emit light upon de-excitation. Electron emission will occur, and atoms and groups of atoms, which may or may not be ionized, will break away from the upper atomic layers at the surface, All these processes can supply information about the surface of the solid.

Spectral analysis of the light emitted by the excited fragments makes chemical analysis of the surface possible. *Fig. 16* gives an example of such an analysis [15]. The method is commonly known as ion-induced light emission (IIL).

SIMS

Mass-spectrometric analysis of the emitted ionized atoms or groups of atoms provides another means of chemically analysing the surface. *Fig. 17* shows the principle of an apparatus used for this technique, known as secondary-ion mass spectrometry (SIMS) [16].

SIMS is essentially a destructive method, since material is sputtered away from the surface during the analysis. In some cases this may be a disadvantage, but it is also possible to turn the sputtering to good advantage. In *dynamic* SIMS, the sputtering is used to obtain a concentration profile, as mentioned at the beginning of this article. This is done by making the primary ion-current density so high (about 10^{-4} A/cm^2) that the surface is eroded away by sputtering in a time that enables the signal to be continuously followed for one or more secondary ions. *Fig. 18* shows an example of some concentration profiles obtained in this way.

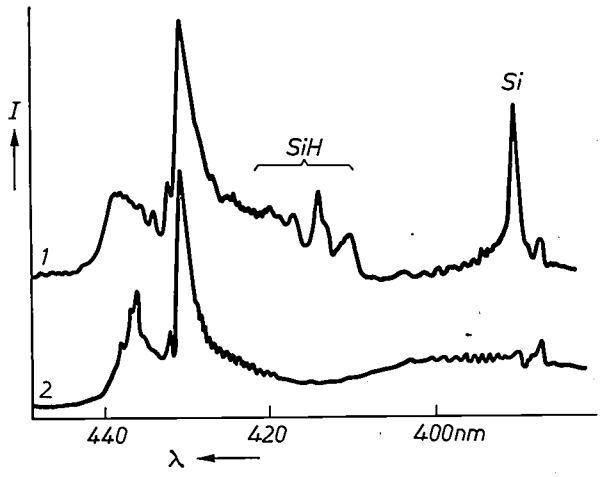


Fig. 16. Intensity I of the light emission from the particles released from a solid surface by ion bombardment, as a function of wavelength λ. Curve *1* gives the emission from a Si surface to which CH_3OH is adsorbed, when bombarded with Kr^+ ions. Curve *2* gives the CH bands in the spectrum of an oxyacetylene flame, recorded with the same spectrometer. The peaks between 420 and 440 nm in curve *1* are thus attributable to CH emission, and so also are the smaller peaks below 400 nm. Comparison with other measurements reveals that the peak at 390 nm originates from Si and the peaks between 410 and 420 nm from SiH.

SIMS analysis not only supplies information about the elements present at the surface of a sample, but also about the chemical compounds in which these elements occur. This is possible because not only atoms but also whole groups of atoms are dislodged from the surface during the processes that give rise to sputtering and ionization. If the experimental conditions are identical, these atomic groups and their relative concentration will also be identical in a particular substance. This enables us to determine the compounds present at the

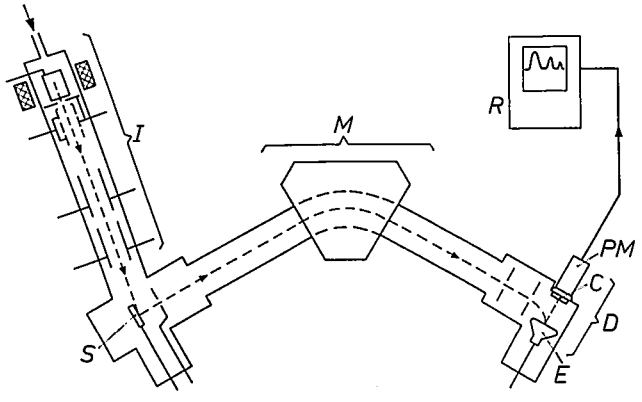


Fig. 17. Diagram of the equipment used for secondary-ion mass spectrometry (SIMS). *I* primary ion source. *S* sample which is bombarded by ions from *I* with an energy of (say) 3 keV. *M* mass spectrometer for analysing the secondary ions (a magnetic 60° single-focusing instrument is drawn here, but in principle any type of spectrometer with a sufficiently high resolution can be used). *D* detector section, where the ions are accelerated to a secondary-emission electrode *E* which is at a high potential. The electrons emitted by this electrode under ion bombardment are detected by a scintillation crystal *C* and a photomultiplier *PM*. The recorder *R* records the secondary ion current from the sample.

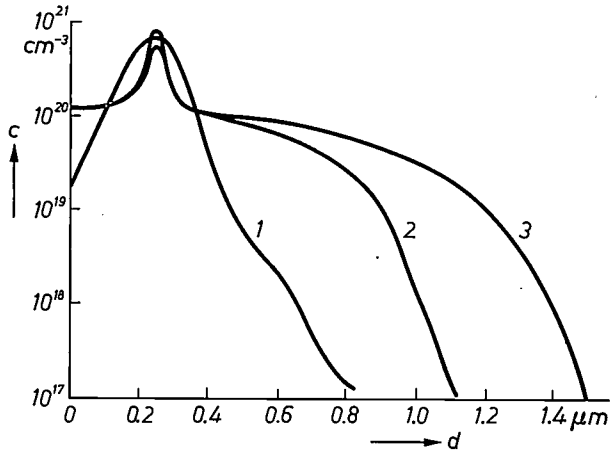


Fig. 18. Dynamic SIMS measurements of the boron distribution in silicon, in which boron ions have been implanted with an energy of 70 keV [17]. The measurements relate to three different samples. Curve *1* gives the boron concentration c as a function of the depth d in a sample after implantation, while curves *2* and *3* give the same concentration for samples that were heated after implantation at 1000 °C for 35 and 100 minutes respectively. It can be seen that the boron ions are taken up in the silicon lattice in two different ways. One group is bound so firmly to its site at a depth of about 0.25 μm that hardly any diffusion occurs even at 1000 °C. A second group is found to diffuse easily through the silicon lattice at the same temperature.

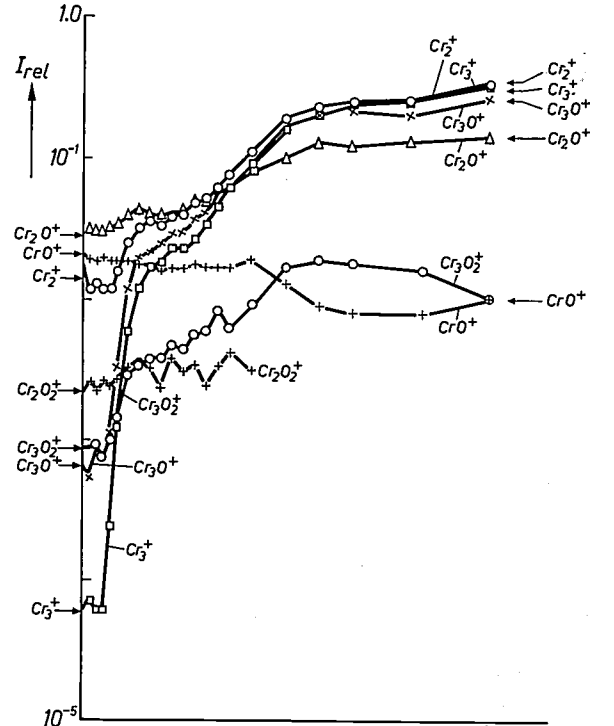


Fig. 19. Dynamic SIMS measurements on an oxidized chromium sample. The different secondary ion currents, I_{rel}, normalized by division by the Cr^+ ion current i_{Cr}, are plotted on a logarithmic scale as a function of the measurement time t, and thus as a function of the depth d below the original sample surface. The 'fingerprint' spectra of pure Cr_2O_3 and lightly oxidized Cr are indicated with arrows at the left and right-hand sides. The oxidized sample surface is seen to have consisted of Cr_2O_3; the measurement three hours later shows that the whole oxide film has been sputtered away down to the chromium.

surface from 'fingerprint' spectra. *Fig. 19* gives an example relating to experiments carried out on a strongly oxidized chromium sample.

If a lower current density is used (about 10^{-9} A/cm^2), the original surface remains virtually intact, only a very minute fraction of the upper atomic layer being removed during a measurement (*static* SIMS) [18]. Provided the pressure in the mass spectrometer is low enough during this measurement, the formation of an adsorbed layer of residual gas atoms will take place more slowly than the sputtering of the surface, so that this contamination does not affect the results of the measurement. If the partial oxygen pressure above the sample is made much higher than the residual gas pressure, the formation of an oxide film can be observed. Under these conditions there will also be an increase in the yield of certain secondary ions, an effect to which we shall return presently.

Finally, it is possible to obtain an actual picture of the distribution of a particular element over the part of the sample surface being investigated. This can be done by scanning the surface with a primary ion beam of about 1 μm diameter and then displaying the signal for the secondary ions of the required element synchronously with the primary beam on a monitor (*microprobe* SIMS), in the same way as in the electron microprobe [19]. Another possibility is to bombard a relatively large part of the surface, with a typical diameter of 300 μm. The secondary ions then give an image of the distribution of the particular element to which the mass spectrometer is tuned, via the electron optics of the mass spectrometer. *Fig. 20* shows an example of the results we have obtained with this technique, described as *imaging* SIMS.

With SIMS, light elements, including hydrogen, can be detected in the sample surface; this is not possible with the electron microprobe.

Quantitative interpretation of the analytical results

The relation between the detected secondary-ion current and the various parameters of the emission process and the instrumentation is given by the expression

$$i_s = i_p S_M{}^+ c_M \eta_M.$$

Here i_s is the secondary and i_p the primary ion current.

[14] A. L. Boers, Ned. T. Vacuümtechniek **11**, 66, 1973.
[15] G. E. Thomas, E. E. de Kluizenaar and M. Beerlage, Chem. Phys. **7**, 303, 1975 (No. 2).
[16] R. F. K. Herzog and F. P. Viehböck, Phys. Rev. **76**, 855, 1949. See also H. W. Werner, Philips tech. Rev. **27**, 344, 1966.
[17] W. K. Hofker, H. W. Werner, D. P. Oosthoek and H. A. M. de Grefte, Appl. Phys. **2**, 265, 1973.
[18] A. Benninghoven, Surface Sci. **35**, 427, 1973.
[19] See for example M. Klerk, The electron microprobe, this issue, page 370.

Fig. 20. Image of a nickel surface recorded with imaging SIMS. In SIMS the crystallites in a sample can be seen because the secondary ion yield varies with the crystal orientation. The light spots in the picture, which in principle seem to indicate a very high nickel content, are in fact nickel oxide particles, which give a higher nickel ion yield than pure nickel.

S_M^+ is the ion yield, the number of secondary ions of the element M present in a concentration c_M at the sample surface, that are emitted per primary ion. The transmission of the mass spectrometer for the element M is η_M.

The ion yield S_M^+ differs from one element to another, and also depends on the environment of the element M (matrix effect). The value of S_M^+ is also dependent on experimental conditions, such as the type and energy of the primary ions and the state of the sample surface. We have found with metals [20], for example, that both oxidation of the surface and the use of oxygen as the primary ion gives an approximately ten-times higher yield of metal ions than bombardment with inert-gas ions.

In principle the ion yield can be determined both by measurement and by calculation. For calculations a theoretical model is needed which, to begin with, must explain the occurrence of ionization and also make the effects encountered accessible to calculation. Various models have been proposed, and some of these will be dealt with here. For practical applications, however, most of the models do not give sufficiently reliable results, so that in general use is made of reference samples or of comparisons with other methods of measurement that are more accessible to calculation.

The lowest concentration c_{Mmin} that can be determined with SIMS depends directly on the lowest current that can be measured. By integrating the current over a long period of time it is possible to measure very small currents, even with a poor signal-to-noise ratio. On the other hand, unduly long measurement times must be avoided to limit the loss of the material, which is particularly important in the analysis of thin layers. For the same reason it is not always possible to lower the minimum detectable concentration by raising the current density in the primary ion beam. It is thus evident that c_{Mmin} is always partly determined by the acceptable material loss. To give some idea of the values involved, we shall mention some details of the analysis of a titanium layer 200 nm thick on an aluminium substrate. At a primary current of 0.45 μA on a surface of 5×10^{-3} cm^2 a secondary current i_{Ti}^+ of 6×10^{-11} A was measured. The rate of erosion of the titanium layer was 0.5 nm/s, so that the entire layer was sputtered away in about 6 minutes. The measurement time for a whole mass spectrum must therefore be relatively short if we are to be able to obtain useful results for the variation of the concentration of certain elements with the depth in the layer. If, however, we only make measurements for a single mass-line, the measurement time for this line can then be longer, to the benefit of the signal-to-noise ratio. If we take a value of 3×10^{-17} A for the detection limit of our current measurement, we can extrapolate from this that 0.5 ppm of titanium must still be detectable.

Ion-emission models

Various models for the emission of secondary ions have been proposed, which were intended to give a quantitative interpretation of the experimental results. All models as yet published are valid only for particular materials or under particular limiting conditions. Unfortunately, a unifying theory capable of explaining all the observations does not yet exist. It is probable that in practice a number of the effects described in the published models will occur simultaneously.

One group of models describes the formation of ions outside the sample. It is assumed that collision processes cause the emission of neutral particles in an excited state, de-excitation can then produce ions, for example by an Auger transition [21]. Although these processes have been intensively studied and are well understood, it is not yet possible to calculate the ion yield for any given primary ion in a given solid. A quantum-mechanical model [22] describing ionization in the sputtering process gives figures, for the ion yield upon Ar^+ ion bombardment of metals, that agree with the experiments to within a factor of 3.

Another group of models explains the increased ionization in oxides as being due to a change in work function. C. A. Andersen [23] uses a thermodynamic model, in which it is assumed that the ions are formed *inside* the solid. In quantitative analyses using SIMS he has obtained good results with this model. In another explanation of the observed enhanced ionization

in oxides it is also assumed that the ions are formed inside the solid, but their formation is explained as being due to the breaking of chemical bonds. One of the considerations underlying this model was our own observation (*fig. 21*) in a SIMS analysis of oxidized copper that the yields of Cu^+ and O^- ions are proportional to one another [20], which warrants the assumption that these ions originate from the same process.

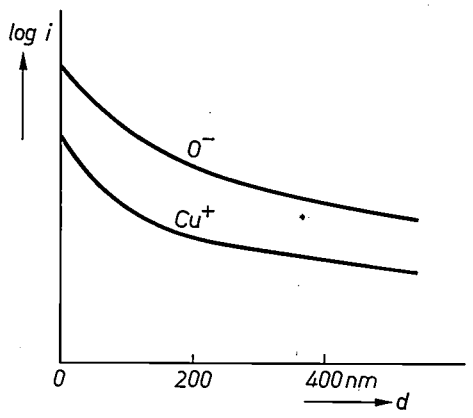


Fig. 21. SIMS investigation of oxidized copper. The secondary ion currents i originating from O^- ions and Cu^+ ions are proportional to one another at all values of the depth d below the surface. This justifies the assumption that both kinds of ion originate from the same process.

Comparison of some aspects of the methods described

All the methods of analysis described here provide particular information about solid surfaces. The detection limit for foreign atoms on the surface varies in the various methods from a few per cent to one ppm of a monoatomic layer.

Depth profiles can be measured by surface erosion with ion bombardment. In SIMS the particles removed from the surface are measured, thus providing information on the bulk concentration, whereas the other methods provide information about the coverage of new surfaces produced. SIMS is the only technique in which sputtering of the surface is essential.

SIMS and Auger electron spectroscopy can provide an actual display of the variation of the investigated property along the surface, with a resolution of a few microns.

Comparison of the various methods discussed in this article also shows that a qualitative analysis of a solid surface can be made with the aid of ion scattering, with SIMS and with Auger electron spectroscopy and ESCA. The last two methods give an analysis of the outer surface layer with a thickness between 1 and 5 nm. Ion scattering and SIMS are confined to the outer atomic layer. SIMS and ESCA not only provide an atomic analysis but also information about the compounds in which the elements occur. In SIMS this is done with 'fingerprint' spectra, in ESCA by using differences in electron binding energies. The concentration profile perpendicular to the surface can be obtained by sputtering the surface layer so as to erode it away layer by layer, something which is always done in SIMS. A quantitative analysis can be obtained with ion scattering, with Auger electron spectroscopy and with ESCA. Until now SIMS has only been able to give a quantitative analysis after calibration by external or internal standards [23].

The geometry of a surface can ideally be studied by means of ion scattering and low-energy electron diffraction (LEED). LEED is particularly useful for studying the symmetry properties of the elementary cell at the surface. The positions of the atoms in the cell can only be determined in certain simple cases. Information about the chemical bonds at the surface can be derived indirectly from observations made with Auger electron spectroscopy and ESCA.

The locations of the electron energy levels at the surface can be deduced from photoemission measurements and from ellipsometric data. As noted earlier, however, ellipsometry is generally confined to measuring the thicknesses of very thin layers.

None of these methods alone, as we have seen, supplies all the desired information. This can usually only be obtained by using several methods in combination. Present trends are therefore towards the development of combined equipment for performing various experiments in one ultra-high-vacuum system, so that the different methods are used for studying *the same* surface.

Summary. Study of the interaction of ions, electrons and photons with a solid surface provides information about the composition of the surface and its structure on an atomic scale. Depending on the method used, the information may relate to the outer atomic layer or to a surface layer with a thickness of 1-5 nm.

A qualitative (usually atomic) and in most cases a quantitative analysis can be made with ion-scattering experiments, secondary-ion mass spectrometry (SIMS) and Auger electron spectroscopy and ESCA (in which electron emission is studied during photon bombardment). Simultaneous sputtering of the surface by ion bombardment (which always takes place in SIMS) yields a concentration profile. Geometric information on an atomic scale can be obtained from ion-scattering experiments and low-energy electron diffraction (LEED). Ellipsometry, a technique normally use for measuring layer thicknesses, supplies information about the position of the electron energy levels at the surface, and similar information is obtained from photoemission measurements.

To obtain a comprehensive picture of the properties of a surface it is always necessary, however, to combine some of the methods discussed. Present trends favour the combination of various techniques in one high-vacuum system, which is the only way in which different observations can be made on the same surface.

[20] H. W. Werner, in: E. L. Grove and A. J. Perkins (ed.), Developments in applied spectroscopy 7A, Plenum Press, New York 1969, p. 239.
[21] R. Castaing and G. Slodzian, C.R. Acad. Sci. Paris 255, 1893, 1962.
[22] J. M. Schroeer, T. N. Rhodin and R. C. Bradley, Surface Sci. 34, 571, 1973.
[23] C. A. Andersen and J. R. Hinthorne, Anal. Chem. 45, 1421, 1973.

Philips tech. Rev. 34, 370-374, 1974, No. 11/12

The electron microprobe

M. Klerk

A strictly localized chemical analysis can often provide extremely useful information for many scientific and technological problems. A localized and usually nondestructive analysis can be made with the electron microprobe, an instrument developed about 25 years ago by R. Castaing and A. Guinier [1]. A sample is bombarded by a fine electron beam that generates, in addition to a continuous X-ray spectrum, a discrete X-ray spectrum characteristic of the chemical elements. The local composition of the sample can be derived from this characteristic radiation.

The electron microprobe has proved to be a useful analytical tool in a wide variety of applications in the electrical-engineering and electronics industries. In the technology of materials and components, for example, it can be used for determining the homogeneity of sintered materials and also for investigating soldered joints and ceramic-to-metal seals. It has its uses in modern semiconductor technology for the analysis of semifinished and finished products, in investigations of production processes and for trouble-shooting in production departments. In metallurgy the microprobe makes it possible to analyse inclusions and identify unknown phases. Mineralogy, biology and archeology are among its other typical fields of application.

The energy of the electrons that bombard the sample must be high enough to excite the characteristic radiation from the constituent elements. There is not usually any advantage in making the beam diameter much smaller than the depth of penetration of the electrons in the sample. To avoid excessive local heating of the sample the beam current must be kept low. Both electrically conducting and nonconducting samples can be investigated, though nonconductors must first be coated with a thin conducting layer (usually a vapour-deposited layer of carbon, aluminium, nickel or copper with a thickness of 10^{-2} μm).

In our instrument the accelerating voltage can be varied between 5 and 50 kV; the beam current, which can be concentrated on a spot with a diameter of about 1 μm, is between 10^{-6} and 10^{-9} A. The volume of material that contributes to the results of the analysis under these conditions is generally about 10 μm^3, or a mass of 10^{-10} g. The smallest detectable amount of an element is about 10^{-14} g.

After a description of the equipment a discussion is given of the methods of solving the problems encountered in evaluating the results of measurements to produce a quantitative analysis. Finally, to illustrate the wide range of applications, some examples are given of analyses carried out with the electron microprobe.

Ir M. Klerk is with Philips Research Laboratories, Eindhoven.

Equipment

The microprobe analytical equipment consists of an electron gun with the optical system required for producing a fine beam, a sample holder, an X-ray spectrometer and a device for displaying the output signal from the spectrometer. *Fig. 1* gives a schematic diagram of the equipment.

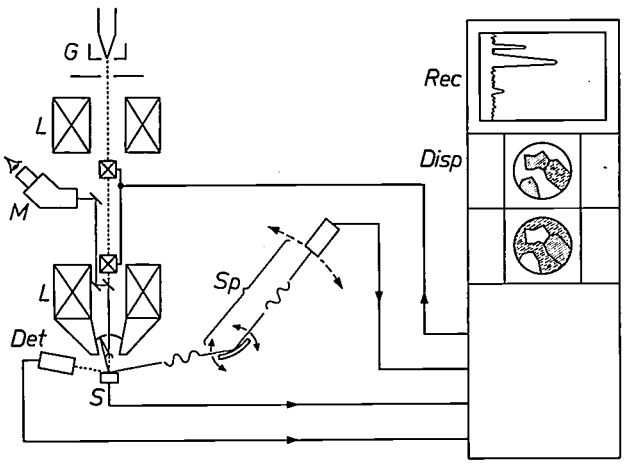


Fig. 1. Diagram of the electron microprobe. *G* triode gun. *L* electromagnetic lenses. *S* sample. *M* microscope. *Sp* crystal spectrometer. *Det* electron detector. *Disp* display tubes. *Rec* recorder.

A triode gun (*G*) delivers an electron beam which is focused by two electromagnetic lenses (*L*) on to the surface of the sample (*S*). The sample, which should be highly polished for a reliable analysis, is placed with a number of reference samples in a special mounting on the specimen stage, which enables the specimen to be displaced in three directions at right angles. An optical microscope (*M*) is used for exact adjustment of the sample height. The microscope has a mirror objective, with a hole at its centre for the electron beam,

so that the sample can be observed during the analysis.

The spectral analysis of the excited radiation is usually carried out with a spectrometer (*Sp*) in which the dispersive element is a crystal. The 'reflection' of X-rays from a crystal is expressed by Bragg's law: $2d \sin \theta = n\lambda$, where d is the distance between the lattice planes in the crystal, θ is the angle of incidence at which the reflection occurs, n is an integer and λ is the X-ray wavelength. For $n = 1, 2, 3, \ldots$ the equation gives the reflections of the 1st, 2nd, 3rd, ... order. Now there can be two wavelengths at which the reflections of different order may coincide. To distinguish between these two reflections, which differ in quantum energy, a proportional gas-flow counter is used as a detector: for each X-ray quantum the counter gives a current pulse whose magnitude is proportional to the quantum energy. The reflections of different order are then separated by a pulse-height discriminator.

Because of the small dimensions and low intensity of the X-ray source it is necessary to use curved crystals. These give the spectrometers better sensitivity than could be achieved with flat crystals, since they enable a greater part of the emitted radiation to be used. It is possible to scan part of the sample surface with the electron beam and to display the detector signal by modulating the intensity of a cathode-ray-tube beam that moves in synchronism with the scanning beam. When the spectrometer is adjusted to the characteristic emission of a particular element, an image of the distribution of that element is obtained in this way. Instead of the X-rays the electrons either back-scattered or absorbed by the sample can also be detected and displayed synchronously. The 'electron images' thus obtained are very similar to photomicrographs.

As well as the equipment described here, which was specially built for the purpose, the scanning electron microscope is also used nowadays for localized chemical microanalysis [2]. Instead of a crystal spectrometer an energy-dispersive solid-state detector is then generally used in combination with a multichannel pulse-height analyser.

Fig. 2 illustrates the variety of information obtainable with the electron microprobe from a sample of a cadmium-bismuth alloy. The information in this case is easily interpreted because the two metals show hardly any mutual solubility and therefore both metals are seen side by side in their pure state. The figure shows two possible modes of scanning. In the first case a small area of the sample surface is scanned, with the spectrometer adjusted to the characteristic radiation of one of the elements in the sample. The resultant image gives the distribution of this element over the surface of the sample. The same scanning method can be used for detecting the back-scattered or absorbed electrons.

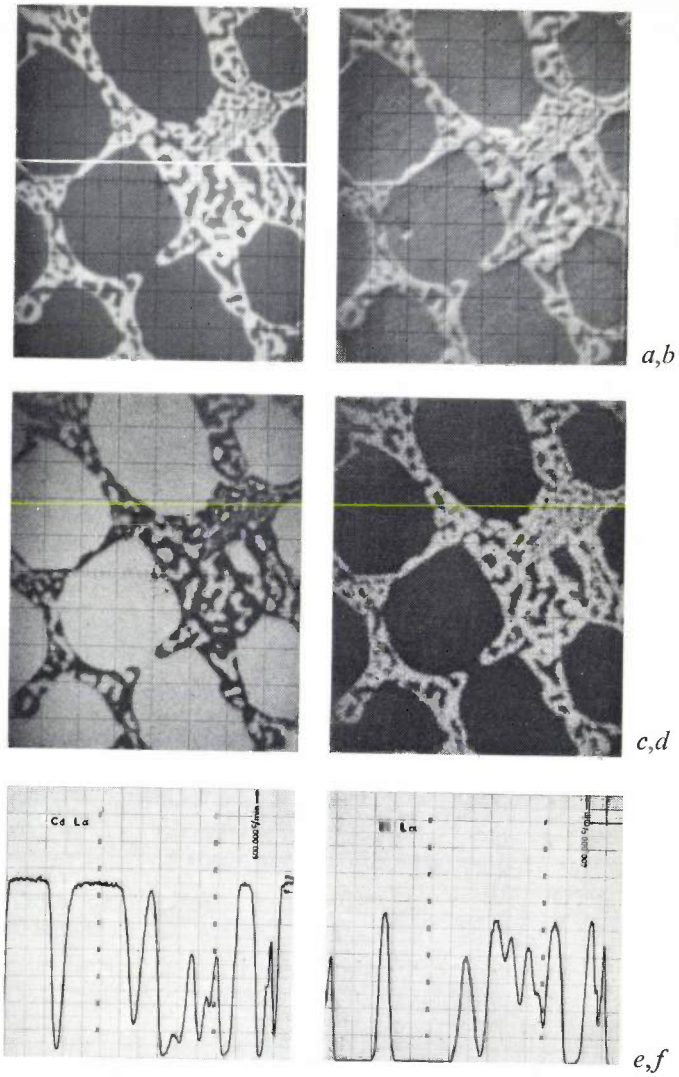


Fig. 2. Images for an alloy consisting of 70 wt% Cd and 30 wt% Bi. The large 'islands' are mainly Cd segregated from the melt; the areas in between consist of the Bi-Cd eutectic. *a*) Image produced by the electrons absorbed in the sample (sample current), represented with 'negative polarity' (black = maximum current). *b*) Image produced by the electrons back-scattered from the sample. The picture shows the relief in the sample surface. *c*) X-ray emission image of $CdL\alpha_1$ radiation. *d*) X-ray emission image of $BiL\alpha_1$ radiation. *e*) The intensity of the $CdL\alpha_1$ radiation, measured along the white line in (*a*). *f*) The intensity of the $BiL\alpha_1$ radiation measured along the same line. Magnification 300×.

In this case the contrast in the image is caused by differences in back-scatter due to differences in chemical composition and possibly to surface irregularities. In the second method a single line on the sample surface is scanned, and again the X-ray emission of one of the elements in the sample is detected. The scanning for this purpose is much slower, which considerably improves the signal-to-noise ratio of the measurement.

[1] R. Castaing and A. Guinier, Proc. 1st Int. Conf. on Electron Microscopy, Delft 1950, p. 60. See also R. Castaing, Adv. X-ray Anal. **4**, 351, 1961.

[2] The next volume of Philips Technical Review will contain an article on the Philips PSEM 500 scanning electron microscope. (*Ed.*)

Processing the results

Qualitative analysis with the electron microprobe is very simple, since it is a simple matter to distinguish the characteristic lines of the chemical elements present in the sample in the X-ray spectrum that it emits.

For quantitative analysis the situation is not quite so simple. The concentrations by weight of the elements in the sample have to be derived from the measured intensities of the strongest characteristic lines. The interaction of the electron beam with the material takes place several microns below the surface of the sample and involves both elastic and inelastic collision and scattering processes. To a first approximation the excitation of a characteristic line by the primary electrons is proportional to the concentration of the element. Closer consideration, however, shows that, owing to various effects, this proportionality no longer holds for the detected intensity of the lines. Because of electron scattering from the atoms of the sample and the energy losses due to collisions, the energy available for excitation also depends to some extent on the other elements present in the sample (*atomic-number effect*). The intensity of a characteristic line can also be increased by contributions originating from fluorescence caused by the shorter-wavelength X-ray radiation of the simultaneously excited characteristic line spectra of other elements present, and from the continuous Bremsstrahlung spectrum (*characteristic* and *continuous fluorescence effect*). Finally, all the radiation excited in the material undergoes absorption before leaving the sample (*absorption effect*).

There are various ways of performing a quantitative analysis. The first and most obvious method is simply to compare the intensity of the emission from the sample with the emission intensity of reference samples of about the same chemical composition. This method has only very limited application, however, because it is difficult and often impossible to make reference samples that are sufficiently homogeneous on a micron scale. At all events, it is very time-consuming, and really homogeneous standards can only be made from elements and simple compounds.

The other and most commonly used methods are based on the determination of the intensity ratios of the characteristic lines emitted by the unknown sample and the identical characteristic lines emitted by reference samples of the pure elements. When deriving the concentrations from these measured relative intensities it is necessary to take the effects mentioned above into account [3]. These effects may sometimes prove to be small or they may largely compensate one another, in which case it can be assumed that the relation between relative intensity and concentration by weight is a linear one, certainly in the case of a short composition range. In other cases, where this is not possible, empirical relations between the relative intensities and concentrations may often be used, enabling a set of calibration curves to be drawn up from only a few reference samples [4].

In the majority of cases, however, a more general method is used for quantitative analysis, which permits the concentrations to be determined by calculating them from the relative intensities. This is done with the aid of analytical expressions, partly theoretical and partly empirical, which allow for the effects mentioned above and give the relative intensity as a function of the concentrations of all the elements in a given sample. Various computer programs are available for performing the rather complicated calculations required [5].

[3] See the review articles by D. M. Poole and P. M. Martin, Metallurg. Rev. **14**, 61, 1969, by D. R. Beaman and J. A. Isasi, Mat. Res. Stand. **11**, No. 11, p. 8 and No. 12, p. 12, 1971, and by S. J. B. Reed, Rev. Phys. Technol. **2**, 92, 1971. A comparable situation is also found in X-ray fluorescence analysis; see for example the article by M. L. Verheijke and A. W. Witmer in this issue, p. 339.

[4] T. O. Ziebold and R. E. Ogilvie, Anal. Chem. **36**, 322, 1964.

[5] D. R. Beaman and J. A. Isasi, Anal. Chem. **42**, 1540, 1970.

Some examples of microprobe analysis

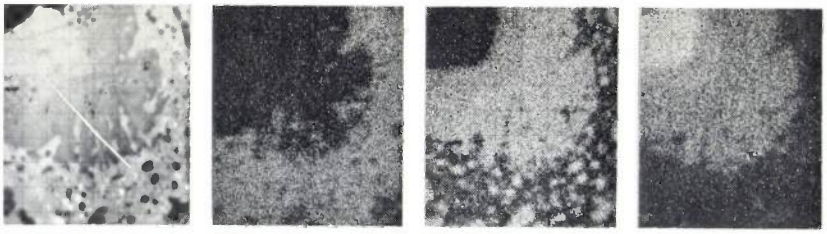
Conductor strips on integrated circuits

absorbed electrons | SiKα | AlKα | OKα

Magnification 125×

We were asked to analyse black spots on aluminium contacts and conductors of an integrated circuit. The resultant microprobe pictures, shown here, in particular that of oxygen, make it clear that the aluminium surface contains oxygen locally; the spots are thus probably caused by oxidation.

The solution of gold wire in tin-lead solder

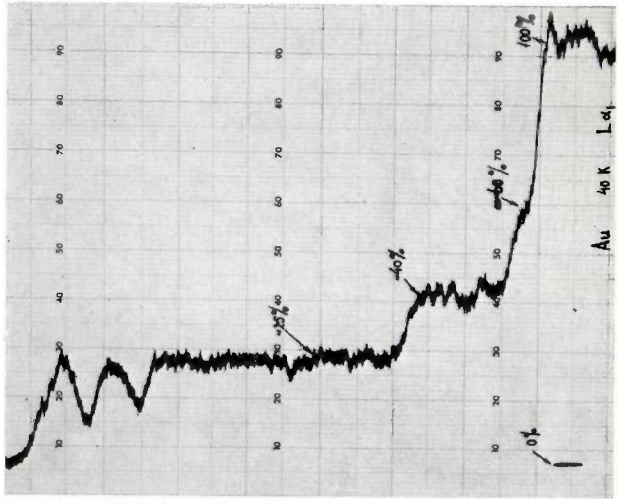


back-scattered electrons | PbLα | SnLα | AuLα
Magnification 250×

An advantage of gold, as compared with copper, is that it can be soft-soldered without a flux. It is advisable, however, to keep the soldering temperature as low as possible so as to avoid excessive solution of gold in the solder. It is even better not to melt the solder at all but to press the gold wires on to a previously tinned metal substrate with a pin whose temperature is just below the melting point of a eutectic tin-lead mixture (183 °C).

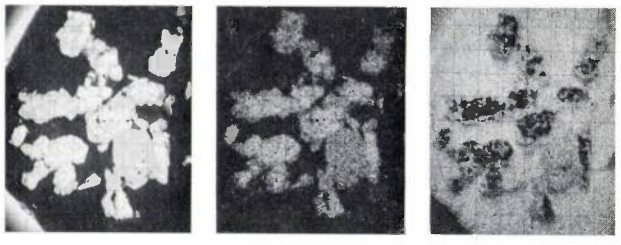
When a 'soldered' joint of this type is heated to 125-175 °C the diameter of the gold wire will gradually decrease. A brittle layer then forms around the wire, which reduces the mechanical strength of the joint. Investigations using the microprobe reveal that this brittle layer contains only gold and tin and no lead. When a line running laterally over a cross-section of the joint is scanned with the probe it is found that the layer consists of a succession of the three intermetallic compounds, AuSn, $AuSn_2$ and $AuSn_4$.

The difficulties caused by the solution of gold in solder can be avoided by using a gold-plated copper wire. The lead-tin solder flows just as well on the gold-plated surface as on pure gold. The amount of gold is too small to give rise to brittle compounds in quantities large enough to cause trouble, and the reactions between copper and solder at about 150 °C are not so rapid as to cause any difficulties.



Variation of the intensity of the $AuL\alpha_1$ emission, measured along the white line in the figure at top left.

Inhomogeneities in a borosilicate glass

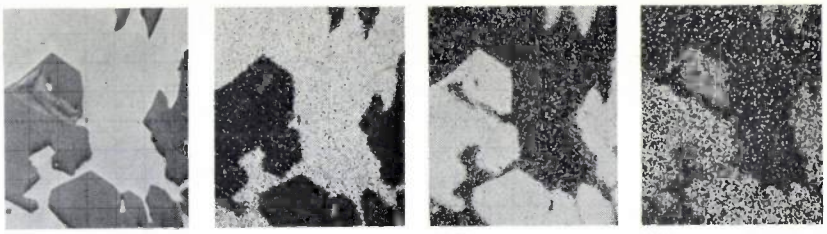


absorbed electrons | ZrLα | SiKα
Magnification 125×

Microprobe investigations on glass samples must be carried out cautiously. Because of the relatively low thermal conduction of the glass the surface temperature can become very high, giving rise to pitting, and electron bombardment also causes local changes in chemical composition due to transport of alkali and alkaline-earth components. Microprobe analyses in this case cannot therefore provide much more than qualitative or semiquantitative results. Subjects for investigation here are local inhomogeneities in the glass, such as stones, inclusions, knots, cords (striae), and also the diffusion profiles that occur when different glasses are fused together. The inhomogeneities may be due to poorly melted components in the batch or to reaction of the glass melt with the furnace lining.

The microprobe images of a stone in a borosilicate glass show that the imperfection consists of two phases. Part of it is ZrO_2 and the rest $ZrSiO_4$. In the cords the concentrations of Zr and Al were found to be locally higher. It was concluded from this that the cords in this case were the remains of corroded 'ZAC' stones from the furnace wall. 'ZAC' stones are pieces of the refractory lining of the furnace wall, cast from the melt, and consist of ZrO_2, Al_2O_3 and SiO_2.

Inclusions in 'Ticonal G'

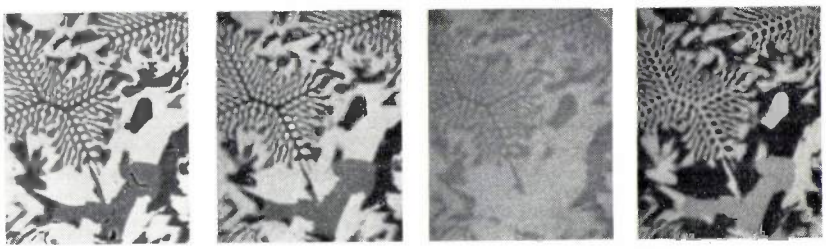


absorbed electrons | CoKα | AlKα | NKα Magnification 125×

'Ticonal G' is an alloy of iron, cobalt, nickel, aluminium and copper. It is a magnetically hard material of high crystalline anisotropy, used for making permanent magnets. Rods of this material, which is rather brittle, are produced in a continuous casting process in which the pellets of alloy are melted in a high-intensity carbon arc in an alumina crucible. Magnets of the required dimensions are obtained from these rods by a grinding process. Some batches of ground 'Ticonal G' showed hair cracks after grinding: when the cracks were examined under the microscope, they turned out to be due to very small inclusions. It was at first assumed that these inclusions consisted of alumina from the crucible. However, a microprobe analysis, resulting in the X-ray emission images shown above, demonstrated that the inclusions were in fact particles of aluminium nitride. These particles were presumably formed as a result of incorrect treatment during melting.

The X-ray image of nitrogen is difficult to obtain. Since nitrogen is a light element, the characteristic radiation is of low energy and low intensity. The radiation is also strongly absorbed by the carbon in the organic materials used in the input windows of the spectrometer and the flow counter.

Analysis of phases in the system Fe-Co-Ti-Ba-O



absorbed electrons | BaLα | TiKα | FeKα Magnification 300×

A composite material consisting of particles exhibiting magnetostriction and particles with piezoelectrical properties may be expected to show magnetoelectric characteristics [6]. A material of this type, in which electric potential differences will arise through the action of a magnetic field, can be obtained by unidirectional solidification of a 'eutectic' mixture of $(CoFe_2O_4)_x(Co_2TiO_4)_{1-x}$, which shows magnetostriction, and the piezoelectric $BaTiO_3$. For reasons not as yet entirely clear, the material shows a stronger magnetoelectric effect when it contains a slight excess of TiO_2 than when it is a pure 'eutectic' [7]. The solid resulting from unidirectional solidification of a melt containing excess TiO_2 has a complicated microstructure. During the solidification, first one, then two and finally three phases segregate simultaneously from the melt. The presence of these phases has been demonstrated by means of X-ray diffraction. They consist of a perovskite phase — $BaTiO_3$ — which is piezoelectric, a spinel phase — $(CoFe_2O_4)_x(Co_2TiO_4)_{1-x}$ — which has magnetostrictive properties, and a magnetoplumbite phase — $BaFe_{12-2y}Co_yTi_yO_{19}$. Analysis using the electron microprobe enabled us to locate the different phases in the sample, which were difficult or impossible to distinguish with conventional metallographic etching techniques. Once the structure of the sample had been discovered by using the microprobe it was possible to develop special etching techniques to make this microstructure visible.

[6] J. van Suchtelen, Philips Res. Repts. **27**, 28, 1972.
[7] J. van den Boomgaard, D. R. Terrell, R. A. J. Born and H. F. J. I. Giller, J. Mat. Sci. **9**, 1705, 1974 (No 10).