

Developments in integrated digital signal processors, and the PCB 5010

J. L. van Meerbergen

Ned. Philips Bedrijven B.V.
PHILIPS RESEARCH LABS.
LIBRARY WY - 1
P.O. Box 80.000
5600 JA EINDHOVEN
THE NETHERLANDS

Since computers first appeared on the scene, more than 40 years ago, their dimensions and energy consumption have rapidly diminished, while their performance and capabilities have steadily increased. Even the vast numbers of calculations per second required for processing digital signals can nowadays be performed on a very small number of chips, or sometimes just one. This is done with a special type of 'computer': the digital signal processor. Some time ago the author of the article below presented a paper^[] on the developments in this field. At Philips, one recent result of these developments is the PCB 5010 integrated digital signal processor. A general picture of these developments and the PCB 5010 itself are the main topics of this article.*

Introduction

For many people, the word 'computer' used to conjure up the standard image of a number of metal cabinets set up in a special room and usually fitted with the well-known magnetic tape units. This picture is changing, however, now that so many of us have home computers and personal computers.

And there seems to be no end to the developments: many users of a Compact Disc player, for example, will not always realize that it contains a special kind of computer for processing the signals. In the near future rather similar devices will also be found in telephones, telephone exchanges, television receivers and all kinds of audio equipment. This will mean the large-scale use of special integrated circuits ('application-specific integrated circuits' or ASICs) designed for digital signal processing. These will include chips designed for one particular application (as in the Compact Disc player, for example) as well as more generally applicable chips, usually known as *digital signal processors* (fig. 1).

The first part of this article gives the general picture of the gradual evolution of the digital signal proces-

Dr Ir J. L. van Meerbergen is with Philips Research Laboratories, Eindhoven

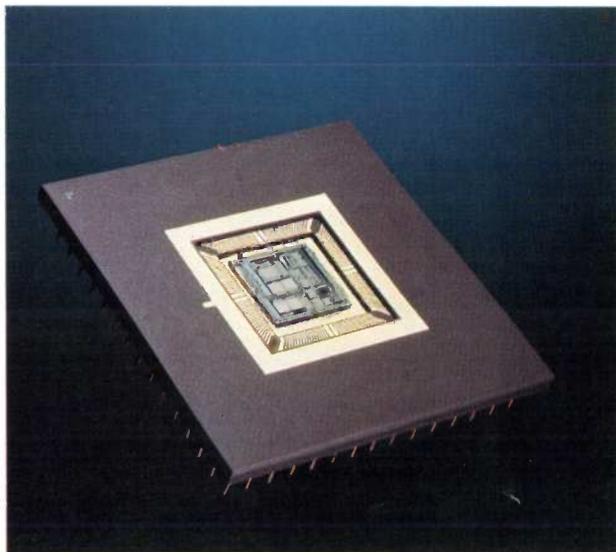


Fig. 1. A digital signal processor is a 'computer on a chip' designed for processing digital signals. These chips contain many tens of thousands of transistors and are designed so that several million typical signal-processing instructions can be executed per second.

[*] J. L. van Meerbergen, Architectures and characteristics of commercially available general-purpose signal processors; paper presented at 'Workshop wave digital filters' IMEC, Louvain, Belgium, 1986.

sor. This is followed by a discussion of the PCB 5010 digital signal processor, now available from Philips, with attention to its architecture and also to programming facilities and supporting accessories. This signal processor is primarily intended for a wide range of applications in telecommunications, audio and speech-processing.

The architecture of computers

Every computing process can be broken down into four elementary operations:

- the input and output of data;
- storage in a memory (of data, intermediate results, final results, and computation procedures or algorithms);
- the execution of the computations;
- the control of the entire process.

This functional division has been of great importance in the design of computers from the earliest days. The

Data, memory addresses and instructions were exchanged in turn via this 'main route'. This is generally referred to as a *von Neumann* architecture (fig. 2). It was the standard computer design for many years, first for the large 'main-frame' computers, and later for minicomputers and microcomputers too. In microcomputers the central processing unit consisted of only a few chips, and was soon to consist of only one — the microprocessor.

Signal processing was performed on main-frame computers right from the start, and efforts were soon made to use microprocessors in the same way. Hopes were expressed that it would be possible to process signals 'in real time', and ultimately on a single chip, so that all the advantages of digital signal processing^[2] would become available for countless applications.

It was soon realized, however, that in some ways the ordinary microprocessor was not so suitable for this purpose and that it was necessary to adapt the de-

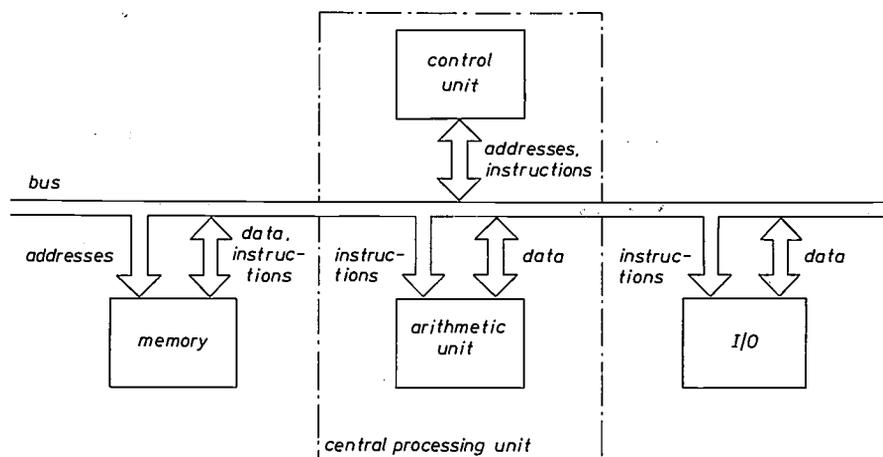


Fig. 2. One of the oldest and best-established forms of computer architecture is the von Neumann architecture shown here. All communication between the main parts of the computer (memory, arithmetic unit, input and output devices I/O, control unit) goes via a single common route: the bus.

four different functions have been performed more or less independently, by:

- input and output (I/O) devices;
- a memory;
- an arithmetic unit;
- a control unit.

The interconnection of the individual devices or units in a specific pattern determined the ultimate architecture of the computer^[1]. Originally a single common provision was made for the interconnections: a *bus*. All communication between the different parts of the computer was made via the bus in successive steps.

sign more specifically to signal processing. The chips that resulted from this development are referred to as digital signal processors. Almost without exception, the first types commercially available had the original von Neumann architecture.

Digital signal processors

So just how do digital signal processors differ from microprocessors? In the first place, a signal processor has to be capable of performing very large numbers of operations per unit time. The exact number depends

directly on the bandwidth of the signals to be processed. At present the most advanced signal processors can handle signals with a bandwidth of up to several tens of kilohertz; integrated signal processors for general applications with video signals are still a thing of the future.

One of the most common digital operations is the multiplication of pairs of numbers from two sequences and the addition of the products, such as

$$y = \sum_{i=1}^N a_i x_i.$$

(Incidentally, this is equivalent to calculating the scalar product of two N -dimensional vectors.) Operations of this type are found in algorithms for filtering, correlation, spectral analysis, etc. To perform them reasonably quickly the arithmetic unit must include a multiplier/accumulator combination or MAC (fig. 3). This can compute one subproduct $a_i x_i$ in the smallest unit of time present in the signal processor (one *clock cycle* or *machine cycle*) and at the same time add the previous subproduct $a_{i-1} x_{i-1}$ to the sum already calculated from all the earlier subproducts. The presence of a MAC is in itself a distinctive difference as compared with the original microprocessors. Much more radical changes in the architecture are required, however, if we are to obtain signal processors that are fast enough; this aspect will be dealt with at some length in the following sections.

The maximum processing rate of a signal processor is also affected by the number of chips used; as the

number of chips diminishes less time will be lost in transferring the signals. As well as an arithmetic unit and a control unit, signal processors must therefore also contain memory facilities. These consist of a ROM (Read-Only-Memory) for storing unchanging quantities, such as the program and constant values, and a RAM (Random-Access Memory) for storing intermediate and final results. The versatility of application is considerably increased if these on-chip memories can be supplemented by external storage as required.

Since the input signals will often be analog in origin and signal processors are digital devices, analog/digital (A/D) conversion will frequently be encountered as a preliminary operation and D/A conversion as a final operation. In one type of signal processor (the Intel 2920) the converters are present on the chip as part of the input/output devices. In general, however, it seems preferable (at least in the present state of the technology) to add the converters to the signal processor as separate components. The specific requirements of individual applications, which can vary considerably, are then more easily taken into account.

The architecture of signal processors

The basic von Neumann architecture has one serious disadvantage: everything happens *consecutively*. Before any one operation is completed, many steps (often very many) have to be completed. For example:

- the location (the 'address') where the next instruction is stored in the memory is determined (e.g. by adding 1 to the previous address);
 - the instruction is read from the memory and transferred to the control unit;
 - the instruction is interpreted ('decoded');
 - the address of data necessary for executing the instruction is sent to the memory;
 - the data is sent from the memory to the arithmetic unit;
 - the arithmetic unit then executes the instruction;
 - the result of the instruction is stored in the memory.
- Then the complete cycle (the *instruction cycle*), which clearly requires more than one *machine cycle*, may be repeated.

The component that restricts the speed most of all is the one most characteristic of the von Neumann structure — the common signal bus, which handles every exchange of information between the various

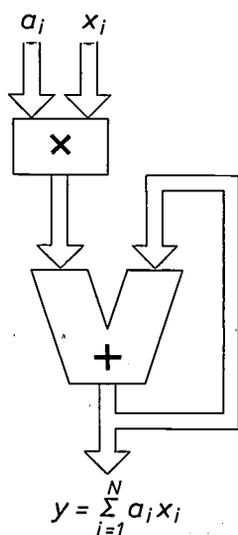


Fig. 3. In digital signal processing it is often necessary to sum N products of the type $a_i x_i$. This takes a disproportionate amount of time in the von Neumann architecture shown in fig. 2. The situation can be improved by providing the arithmetic unit with a multiplier/accumulator combination as shown here.

[1] The combination of the arithmetic unit and the control unit is often called the central processing unit.

[2] J. B. H. Peek, Digital signal processing—growth of a technology, 103-109, in the special issue 'Digital signal processing I, background', Philips Tech. Rev. 42, 101-144, 1985.

parts of the processor. It acts as a bottleneck. For higher speeds it is necessary to change to a 'non-von Neumann' architecture (sometimes just called a 'non-von' architecture).

The most common alternative is the Harvard architecture, in which data and instructions are stored in separate memories and which therefore has to have separate connections for data and control information. This gives the architecture shown in *fig. 4*, where the signal processor is divided into two parts, called the controller and the data path. The exact nature of the connections *between* these two parts is mostly of less importance.

The architecture of the PCB 5010

The design of our PCB 5010 signal processor is based on the Harvard architecture of *fig. 4*. It has been modified in some essential aspects, however, to obtain sufficient versatility and signal-processing capacity.

One of the main modifications is the duplication of the data bus, resulting in an X bus and a Y bus, each 16 bits wide^[3]. The reason for this is that most signal-processing operations have two operands and can only be performed efficiently with a duplicated data bus. There is also the advantage that complex numbers can be more easily manipulated with a duplicated bus. As a direct consequence of this duplication, the RAM for the data is divided into two parts. A separate ROM is also available for data.

As we have seen, every signal processor has to have a multiplier/accumulator combination *MAC* to reach the speed required in the many 'vector-like' operations that arise. High-speed processing is also required for other types of operation, such as logic AND, NOT, OR etc., for operations on absolute values and for operations on individual bits ('masks'). The PCB 5010 has a separate arithmetic and logic unit *ALU* for these activities.

Finally, to facilitate data input and output our signal processor has two serial input devices and two serial output devices, as well as a combined parallel input/output device. This brings us to the block diagram^{[4]-[6]} of the PCB 5010 in *fig. 5*.

A computer drawing of the actual plan of the PCB 5010 is shown in *fig. 6*; the numbers 1 to 5 refer to the main components in the previous figure. There are some 135 000 transistors in all in this drawing, yet this IC only occupies an area of 61 mm² when fabricated in 1.5- μ m CMOS technology.

In the following sections of this article we shall take a closer look at the structure of the controller and of the two principal components of the data path: the

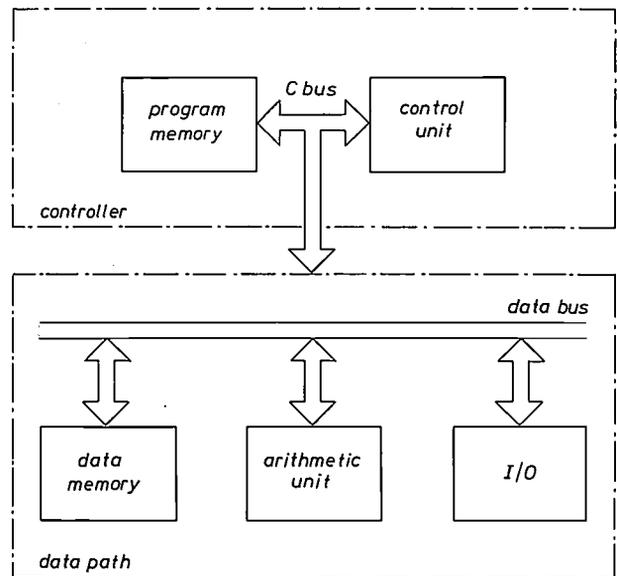


Fig. 4. A modern alternative to the von Neumann architecture is the Harvard architecture shown here. This has separate memories for the program and the data, and separate connections for the control information and the data. These connections are called the control bus (or C bus) and the data bus. The total architecture can now be divided into two parts, called the controller and the data path.

data memory and the arithmetic unit. We shall see that in each of these components every effort is made to decentralize as many activities as possible to prevent bottlenecks, e.g. by providing a separate address-computation unit for each memory.

Besides the architecture, various other aspects are important in assessing the performance of a signal processor. They include:

- the time (in seconds and in number of machine cycles) required for executing an instruction;
- the number of (sub)operations that can be performed simultaneously;
- facilities for interaction with the outside world;
- the possible degree of overlap in time of executions of successive instructions in different parts of the signal processor ('pipelining').

These all depend greatly on the way in which the signal processor can be programmed (the 'microcode'). We shall return to this point later.

The ultimate critical factor in comparing signal processors is the time required for performing a num-

^[3] F. P. J. M. Welten *et al.*, A 2- μ m CMOS 10-MHz microprogrammable signal processing core with an on-chip multiport memory bank, *IEEE J. SC-20*, 754-760, 1985.

^[4] J. L. van Meerbergen *et al.*, A 2- μ m CMOS 8-MIPS digital signal processor with parallel processing capability, *Int. Solid State Circ. Conf. (ISSCC)*, Digest of technical papers, Anaheim, Cal. 1986.

^[5] F. J. van Wijk *et al.*, A 2 μ m CMOS 8-MIPS digital signal processor with parallel processing capability, *IEEE J. SC-21*, 750-765, 1986.

^[6] Introducing the PCB 5010/PCB 5011 programmable DSPs, Philips Electronic Components and Materials Division, Eindhoven 1986 (16 pages).

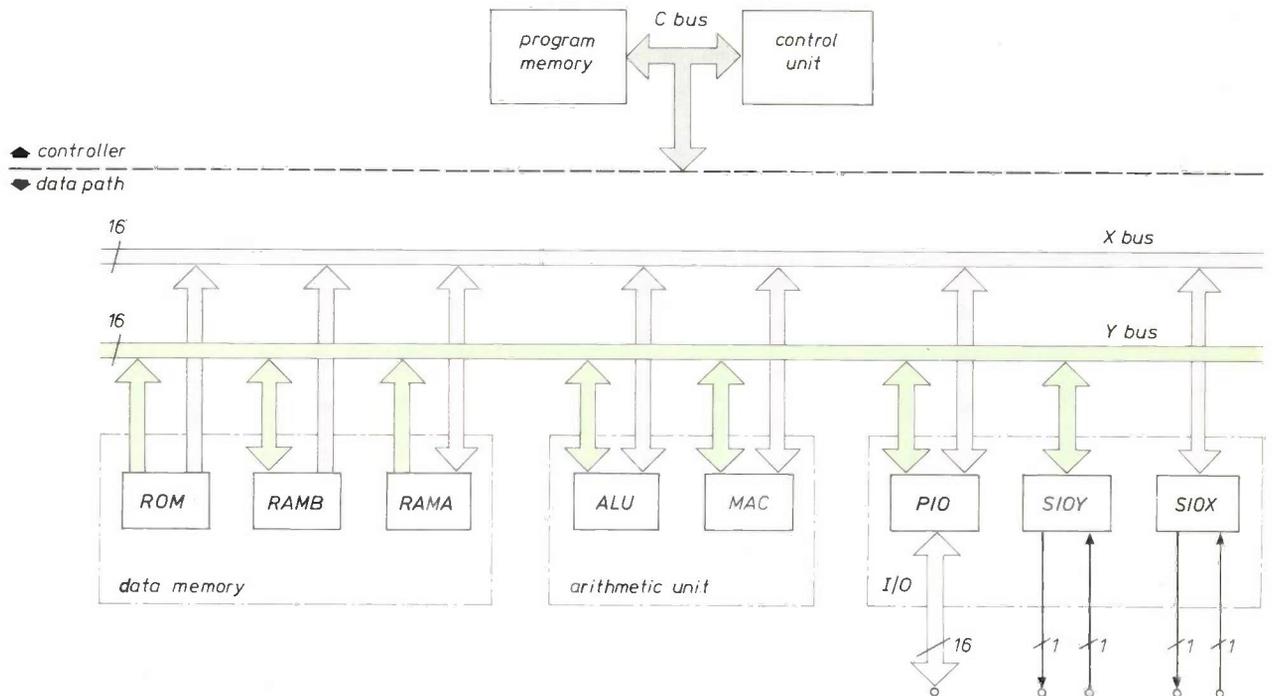


Fig. 5. The PCB 5010 digital signal processor has a Harvard architecture with two independent 16-bit data buses: the X bus and the Y bus. The data memory consists of three parts: a ROM (read-only memory) for unchanging data and two RAMs (random-access memories) RAMA and RAMB. The arithmetic unit consists of two parts: a multiplier/accumulator combination MAC and an arithmetic and logic unit ALU. The input and output devices I/O consist of a parallel unit PIO, which has a 16-bit connection to the outside world, and two serial units SIOX and SIOY, each with a 1-bit input and a 1-bit output.

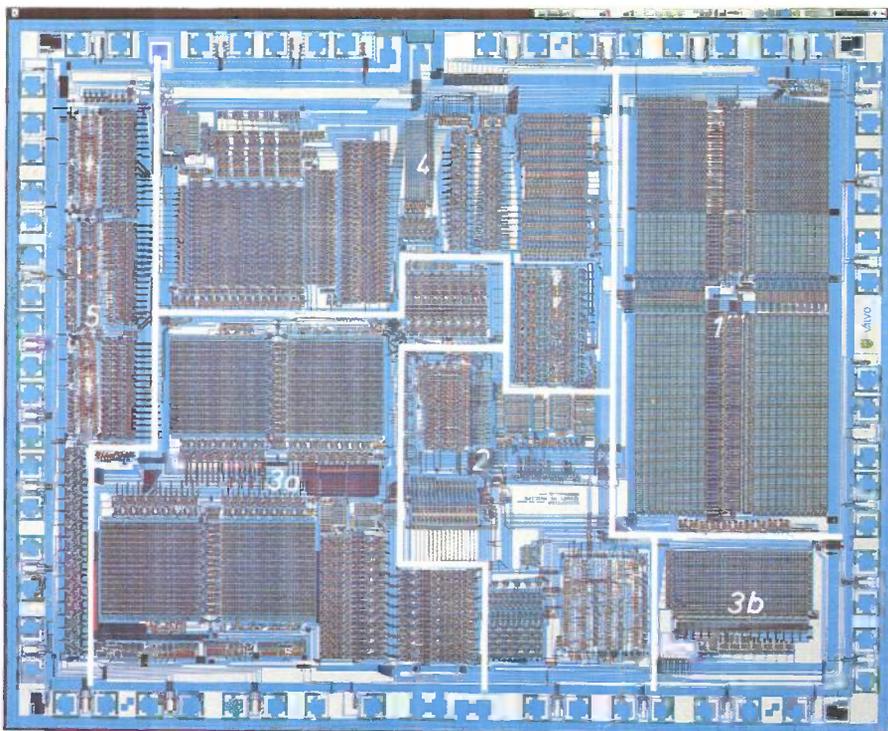


Fig. 6. Computer drawing of the integrated digital signal processor PCB 5010. The most important components are: 1 program memory; 2 control unit; 3a, 3b data memory; 4 arithmetic unit; 5 input and output devices I/O. The complete circuit contains some 135 000 transistors and is fabricated in 1.5- μm CMOS technology. It occupies an area of 61 mm².

ber of standard operations at a given accuracy. Typical operations might be a 128-point FFT ('Fast Fourier Transform'), a complex multiplication or a particular elementary filtering operation. These operations are often called 'benchmarks'.

The arithmetic unit

The multiplier/accumulator combination MAC

Fig. 7. shows a block diagram of the multiplier/accumulator with its various supporting devices. The operation of multiplier *MPY* is based on the 'modified

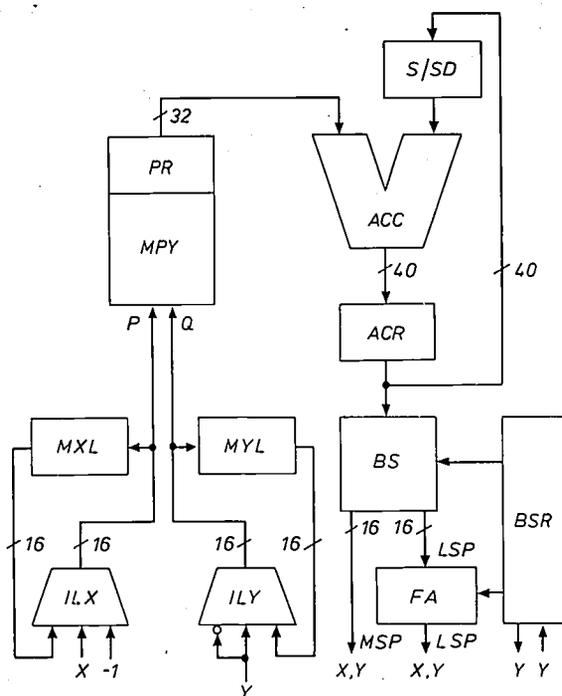


Fig. 7. Block diagram of the multiplier/accumulator combination *MAC* of the PCB5010. The number of operations that can be performed in each machine cycle has been greatly increased by the addition of extra registers and selectors. The operations required are specified by a 7-bit code via the C bus (not shown explicitly). The numbers beside the oblique strokes in the connections indicate how many bits are transferred in parallel. The significance of the designations is:

<i>ACC</i>	accumulator
<i>ACR</i>	accumulator register
<i>BS</i>	barrel shifter
<i>BSR</i>	barrel-shift register
<i>FA</i>	format adjuster
<i>ILX</i>	X-input selector
<i>ILY</i>	Y-input selector
<i>LSP</i>	least-significant part
<i>MPY</i>	multiplier
<i>MSP</i>	most-significant part
<i>MXL</i>	X latch
<i>MYL</i>	Y latch
<i>P</i>	operand 1
<i>PR</i>	product register
<i>Q</i>	operand 2
<i>S/SD</i>	sign/scale-down block
<i>X</i>	16-bit connection to the X bus
<i>Y</i>	16-bit connection to the Y bus

Booth's algorithm'^[7] and in one machine cycle it can compute the product of two 16-bit words *P* and *Q*. The product is stored temporarily as a 32-bit word in the product register *PR*. At the same time the accumulator *ACC* can add the previous contents of *PR* to the contents of the accumulator register *ACR*. The contents of *ACR* are then multiplied, under the control of the block *S/SD*, either by 1, -1, -2^{-15} , 2^{-15} or by 0. The output of the accumulator has a width of 40 bits, so that even if large numbers of products are added together, there are no overflow errors.

The 'barrel shifter', *BS*, the corresponding barrel-shift register *BSR* and the format adjuster *FA* derive one group or two groups of 16 bits from the 40-bit contents of *ACR* to form the output signal of the *MAC* unit.

The operands *P* and *Q* presented to *MPY* may come directly from the X bus and the Y bus via the X and Y input selectors *ILX* and *ILY*. It is also possible to select the *previous* X and Y information, which is automatically stored in the latches *MXL* and *MYL*. In addition, *P* can take the value -1, and *Q* the logically inverted value of the current Y information.

Certain special occurrences, such as overflow in *ACC*, are reported directly to the control unit by means of a 'flag' or 'flag signal'.

Multiplication with greater precision

Blocks *S/SD*, *BS*, *BSR* and *FA* are also important because they permit calculations to be made at a greater precision than 16 bits, though at the expense of more processing time. For example, the product of a 46-bit operand and a 31-bit operand can be computed in 7 machine cycles at most^[8]. This is done in much the same way as multiplying two large numbers together conventionally: subproducts are determined first, then shifted appropriately and added.

The arithmetic and logic unit ALU

A block diagram of the arithmetic and logic unit *ALU* of the PCB 5010 is shown in fig. 8. Grouped around the logic unit *LU*, which can perform operations on a single operand *A* or on two operands *A* and *B*, there are a number of registers and other supporting devices.

A total of 31 different arithmetic and logic operations can be executed, including addition, subtraction, absolute-value determination, AND, OR, NOT and shift operations. The selectors *ILA* and *ILB* permit either the current information on the X and Y buses to be used for *A* and *B*, or the previous information, which is always automatically stored in the latches *AAL* and *ABL*.

At the output of *LU* a set of 15 registers are available for temporary storage. During each machine cycle the present result of *LU* can be stored in one of these registers, and the contents of any other two registers can be read out via the *X* bus and the *Y* bus.

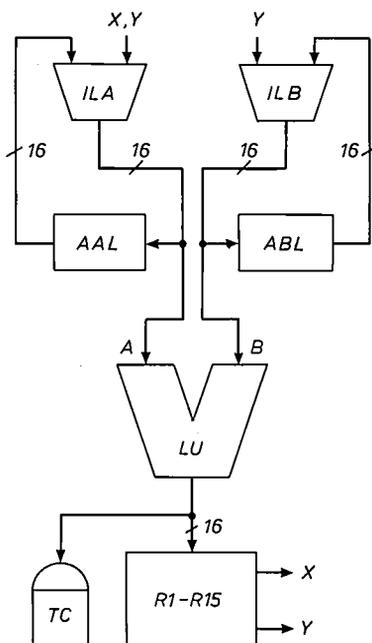


Fig. 8. Block diagram of the arithmetic and logic unit *ALU* of the PCB 5010. As well as two selectors and two latches, this unit contains a file of 15 independent 16-bit registers. The operations required from *ALU* are specified in each machine cycle by a 7-bit code via the *C* bus. Also, 4 bits are reserved on the *C* bus for selecting one of the registers from the register file. The significance of the designations is:

- A* operand 1
- AAL* A latch
- ABL* B latch
- B* operand 2
- ILA* A-input selector
- ILB* B-input selector
- LU* logic unit
- R1-R15* register file
- TC* trash can
- X* 16-bit connection to the *X* bus
- Y* 16-bit connection to the *Y* bus

The arithmetic and logic unit *ALU* can send status information directly to the control unit of the signal processor by means of flags, (e.g. about overflow or the sign of the computed result). If just this status information is to be stored, the rest of the computed result is consigned to the 'trash can' *TC*.

If required, *ALU* can perform computations to a higher precision than 16 bits, but again it will take longer for the processing.

The units *MAC* and *ALU* are to some extent complementary. *MAC*, for instance, is designed for processing 16-bit words (and hence for vector operations), while *ALU* is particularly suited for processing individual bits. There are seven types of instructions for *MAC*, all concerned with multiplication and addition, while *ALU* has 31, some very different from the others. Finally, in *MAC* there is always room for intermediate and final results with a maximum length of 40 bits; in *ALU* the length is always 16 bits, unless special arrangements are made. These differences are summarized in *Table I*.

Table I. Comparison of some complementary features of the multiplier/accumulator combination *MAC* and the arithmetic and logic unit *ALU*.

<i>MAC</i>	<i>ALU</i>
Designed for vector operations (16-bit word level)	designed for other operations (bit level)
7 slightly different instructions (all connected with adding and multiplying)	31 sometimes very different instructions (such as AND, OR, NOT, EXOR, shifting, addition, subtraction, incrementing)
Gives 40-bit intermediate result and 40-bit final result (can be extended)	gives 16-bit result (can be extended)

The data memory

As already mentioned, three memories are provided for storing all the various data required during the operation of the signal processor. One memory with 512 locations, each of 16 bits, stores unchanging data, such as constants and filter coefficients, and is therefore a ROM. This data is entered into the memory once only, during manufacture. The other two are RAMs; these store intermediate or final results. They each have 128 memory locations of 16 bits (*fig. 9*). Each of the three memories has its own output-data register (*DRR*, *DRB* and *DRA*).

To keep the capacity of the arithmetic unit and the *X* and *Y* buses of the signal processor free as far as possible for the actual signal-processing operations, each of the three memories has its own address-computation unit, or ACU, denoted by *ACUR*, *ACUB* and *ACUA*. During each machine cycle a new mem-

[7] A very good short description of this algorithm is given in: L. P. Rubinfield, A proof of the modified Booth's algorithm for multiplication, IEEE Trans. C-24, 1014-1015, 1975.

[8] The length of the operands is not quite an integer multiple of 16, because one bit in each group of 16 is always reserved as a sign bit, and the length can only increase effectively by multiples of 15, thus: 16, 31, 46, ...

ory address can be computed in an ACU from the old address and some additional data. For example, a fixed number can be added to the previous address. If this is a 'modulo- N ' addition, the data memory is scanned repeatedly in a fixed pattern, which is very useful for repetitive look-up of filter coefficients. Another possible operation is the reversal of the order of the address bits, which facilitates the calculation of Fast Fourier Transforms.

In both *RAMA* and *RAMB* the computed addresses are longer than is necessary for addressing 128 differ-

ent locations. As *ACUB* supplies an 8-bit address, a later version of this signal processor can thus have a larger *RAMB* with 256 locations. *ACUA* supplies a 12-bit address, which is also available at the connector pins of the signal processor for addressing an external memory. It can be extended by 4 bits from a page register *PG* to form a 16-bit address, so that an external data memory with 64k (= 65 536) locations can be used.

The controller

The controller is the 'brain' that directs the entire operation of the signal processor. It determines what happens in the signal processor in each individual machine cycle. The program memory is of major importance here. It stores all the possible instructions for all the components of the signal processor. These instructions have a fixed length of 40 bits. The program memory of the PCB5010 can contain 1024 such instructions (992 in a ROM and 32 in a RAM). The instructions stored in the ROM are entered permanently when the signal processor is manufactured; the instructions in the RAM can be changed at any time. During each machine cycle a 40-bit instruction is entered into the instruction register *IR* (fig. 10), and the various sub-instructions are then distributed over the entire chip. A part of some instructions goes straight to the X bus and the Y bus ('load immediate', see the section on the microcode); the rest of the information in the instructions goes to the various parts of the signal processor via connections that we shall not consider further here (the C bus). At the same time the program counter *PC* determines the next address for the program memory.

For certain frequently occurring types of program special provisions have been made. A particular instruction may have to be repeated N times, and a separate instruction-repeat register (*RPR*) is available for such operations. While this is in use, the contents of the program counter *PC* remain unchanged. Other regularly occurring events include interruptions and the execution of subroutines. The current program is then stopped for a moment, and this is noted in a 'stack register', so that the program can be resumed later. The stack register has five levels, but if required it can be extended by a part of one of the data memories described in the previous section. The stack register enables other subroutines and interruptions to be processed inside a subroutine.

Since the controller represents the 'nerve centre' of the signal processor, it has many connections to the outside world and other parts of the chip. We have already encountered several examples, such as the capa-

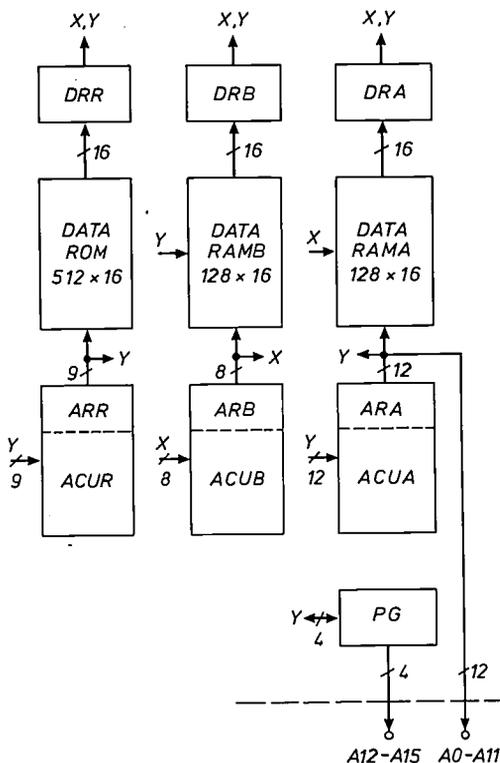


Fig. 9. Block diagram of the three data memories of the PCB 5010. Each memory has its own address-computation unit and its own output-data register. The 12-bit address of one of the two RAMs is also available externally. If this address is extended by an extra 4 bits from a page register, an external data memory with 64k (= 2^{16}) memory locations can be used. In each machine cycle for each of the address-computation units three bits of information are reserved on the C bus. The significance of the designations is:

<i>ACUA</i>	address-computation unit A
<i>ACUB</i>	address-computation unit B
<i>ACUR</i>	address-computation unit R
<i>ARA</i>	address register A
<i>ARB</i>	address register B
<i>ARR</i>	address register R
<i>A0-A15</i>	address bits 0-15
<i>DRA</i>	output-data register A
<i>DRB</i>	output-data register B
<i>DRR</i>	output-data register R
<i>PG</i>	page register
<i>RAMA</i>	random-access memory A
<i>RAMB</i>	random-access memory B
<i>ROM</i>	read-only memory
<i>X</i>	16-bit connection to the X bus
<i>Y</i>	16-bit connection to the Y bus

bility for external interruption of a current program. Clock-synchronization signals and any reset signals also reach the controller from outside. Status information about the various components of the signal

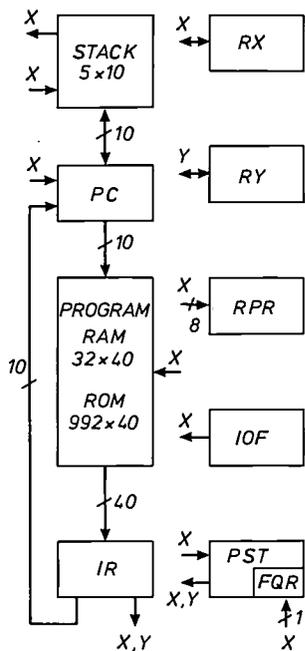
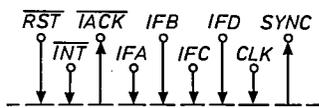


Fig. 10. Block diagram of the controller of the PCB 5010, consisting of the program memory and the control unit. The program memory is mostly read-only, with a small random-access section. The memory address is supplied by the program counter. The contents of the program memory are written to the 40-bit instruction register, and from there the information is distributed via the C bus and some of it possibly also via the X bus and the Y bus. The control unit consists mainly of a number of registers that serve a variety of purposes. The arrows at the top indicate that the control unit can exchange certain information with the outside world. The significance of the designations is:

- CLK clock
- FQR mode bit (P mode/NP mode)
- IACK acknowledgment
- IFA user flag A
- IFB user flag B
- IFC user flag C
- IFD user flag D
- INT interrupt signal
- IOF input/output status and user flag register
- IR instruction register
- PC program counter
- PST processor-status register
- RAM random-access memory
- ROM read-only memory
- RPR instruction-repeat register
- RST reset signal
- RX X register
- RY Y register
- STACK stack register
- SYNC synchronization signal
- X 16-bit connection to the X bus
- Y 16-bit connection to the Y bus

processor is always stored in the 16-bit processor-status register *PST* in the form of 1-bit flag signals. Similarly the register *IOF* contains input/output flags that indicate the status of the input and output circuits of the processor, and four user flags that originate outside the chip. Finally, the controller contains two registers, *RX* and *RY*, that save the signals on the X bus and the Y bus if there is an interruption.

General diagram

A block diagram giving a general picture of the component parts of the PCB 5010 signal processor discussed above is shown in *fig. 11*. The numbers beside the interconnections indicate the number of parallel bits. The symbols *X* and *Y* indicate a direct connection to the X bus and the Y bus. Some connections to the outside world are also indicated (the chip is in an encapsulation with 68 connector pins). The internal connections for control purposes (the C bus) are not shown, however.

Programming

The architecture of signal processors has been discussed at some length above, because it very largely determines the *theoretical* processing capabilities of the signal processor, such as the maximum available degree of parallelism. The *actual* processing capacity is also very dependent, however, on the ‘programmability’ of the processor. This can be deduced from the structure and diversity of the instructions (the ‘microcode’) that can be used to program the chip. Another extremely important point here is the amount of effort required from the user to translate a particular required function into a processing program (an algorithm) and hence into a sequence of basic instructions. This is mainly determined by two factors:

- the ‘transparency’ of the microcode;
- the facilities available for creating, testing and correcting a processing program.

We shall now look first at the microcode itself, and then at the available facilities.

The microcode

The PCB 5010 works with 40-bit instructions and a machine cycle time of 125 ns. In the *pipeline mode* (the P mode) each machine cycle corresponds to a single instruction; in the *non-pipeline mode* (the NP mode) the instructions follow one another at regular intervals of two machine cycles (we shall return to this point later). Each instruction represents one or more basic operations. In one machine cycle, for example,

the entire data stream of a program can be selected and then the corresponding address computations can be determined.

Accessories

Besides the PCB 5010, there is a very similar processor, the PCB 5011. This has no program memory, however, and no on-chip data ROM; the chip has connector pins for these, so that external memories can be connected to it. The PCB 5011 contains 70 000 transistors, while the PCB 5010 has about 135 000. Fig. 13 is a photograph of the PCB 5011. The left-hand two-thirds of fig. 6 can clearly be identified here. The large number of external connections is also immedi-

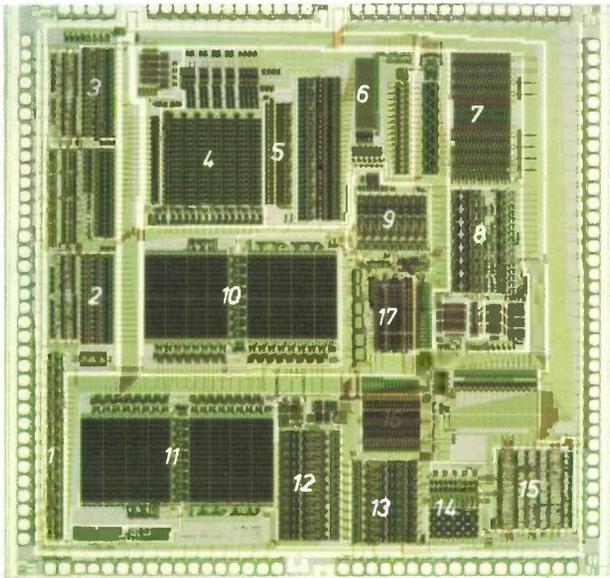


Fig. 13. Photograph of the PCB 5011 digital signal processor. This processor is identical to the PCB 5010 in fig. 6 and fig. 11, except that it does not have the program memory and the data memory ROM. The connections for these go to connector pins, so that external memories can be used instead. The following components are indicated by numbers (the abbreviations are the same as in the previous figures): 1 PI and PO; 2 SIX and SOX; 3 SIY and SOY; 4 multiplier; 5 accumulator; 6 barrel shifter; 7 register file; 8 logic unit; 9 ACUB; 10 RAMB; 11 RAMA; 12 ACUA; 13 ACUR; 14 STACK and PC; 15, 16 various components of the control unit; 17 PST.

ately obvious: the PCB 5011 has 144, the PCB 5010 'only' 68. (The integrated circuit shown in fig. 1, incidentally, is also of type PCB 5011.)

The PCB 5011 can be used in the design or development phase of a system (before any final decision has been made about the contents of the ROM in the PCB 5010), in applications where it is not worth

making a special version of the PCB 5010 with ROMs specified by the user, and in applications where a very large program memory is required.

To facilitate the use of the PCB 5010/PCB 5011 special software has been written in the programming language PASCAL. This can be used on several widely used computers (VAX, IBM-PC). In the first place there is a *simulator program*, which can simulate the entire operation of a signal processor programmed for a specific application. There is also an *assembly program* that makes it unnecessary to specify the contents of the instructions bit by bit, requiring only symbolic indications that can be handled more readily (i.e. groups of letters — 'mnemonics' — that look like abbreviations). In this program these are automatically translated into bit sequences. The program also contains a *macro library*, in which frequently occurring algorithms, such as certain kinds of filtering and FFT operations, are stored in 'macrocode' as ready-to-use subroutines for the signal processor.

To test a system in which the PCB 5010/PCB 5011 is used under realistic conditions, e.g. in real time, the 'Stand-alone Debug System' (SDS) can be used. The SDS is an *emulator*, i.e. a device that functions in exactly the same way as a later definitive version of the PCB 5010, but also has a variety of facilities for interrupting a program being run in the signal processor at any moment and for investigating the internal status of the signal processor at that moment. Program modifications are also easily made.

Finally, there is a *prototype board* containing the PCB 5011 and all the external memories and circuits required for loading these memories. This board can be used, for example, for making a prototype of a system that will later include one or more PCB 5010 chips.

Applications

The PCB 5010 has its greatest signal-processing capacity when it is used in the pipeline mode: in pipelining a new operation starts while the last part of the previous operation is still being performed at another location on the chip during the same machine cycle. This is done in product accumulation, for example. A single product accumulation really takes two machine cycles, but by pipelining the multiplication part and the addition part the effective duration is only about one machine cycle in long sequences of product accumulations. Programming in the pipeline mode is rather more difficult than in the non-pipeline mode. It is therefore possible to select one of the two modes and even to switch from one to the other within the same signal-processing program.

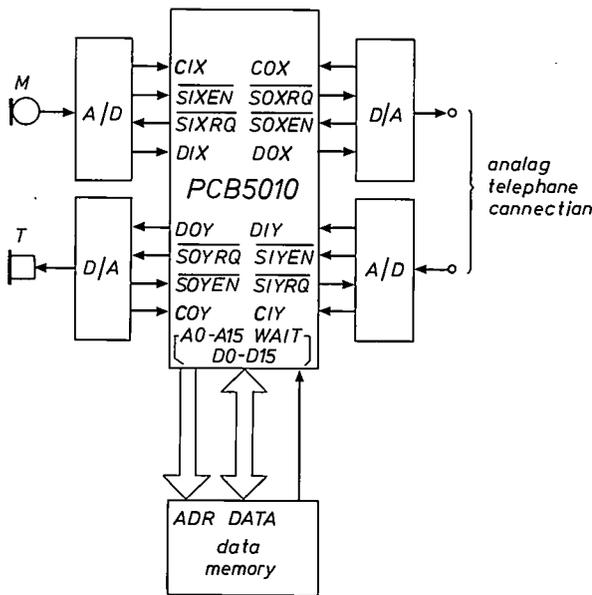


Fig. 14. Example of the application of the PCB 5010 in telecommunications. The figure shows how the signal processor can be used in an analog telephone connection: after addition of a telephone *T*, a microphone *M* and A/D and D/A converters, a single PCB 5010 can perform all the processing operations required for transmitting and receiving. It can also be seen how the internal data memory RAMA can be replaced by an external memory.

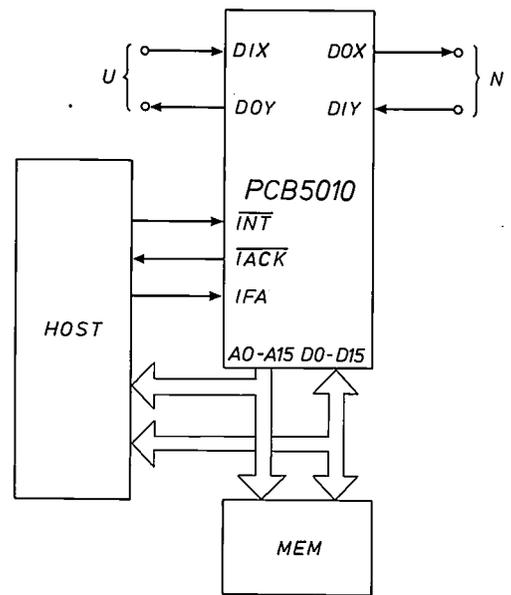


Fig. 15. A PCB 5010 can be controlled from an external microprocessor *HOST* in combination with an external data memory *MEM*. The example shown here relates to a terminal in a telecommunication system; the PCB5010 links the user *U* to the rest of the network *N*.

The PCB 5010 can be used in a variety of system configurations; in a minimum configuration only A/D and D/A converters have to be added. In many communication applications incoming and outgoing signals can even be processed effectively simultaneously (fig. 14). It is also possible to use an external microprocessor to control the signal processor; this makes it even more versatile (fig. 15). Also, since the PCB 5010 chips have extensive input/output facilities, they can easily be combined to form a multiprocessor system suitable for more demanding applications (fig. 16).

Just how powerful the PCB 5010/PCB 5011 signal processors are can be seen most clearly from the time required to execute a number of characteristic processing operations. A summary of these is given in Table II. Unless otherwise stated, the information in this Table relates to 16-bit quantities, for both signal samples and filter coefficients. In the examples relating to the Fast Fourier Transform (FFT) 'looped code' processing programs were used. This reduces the number of steps in the program to only about 40. It is also possible to halve the time required by using 'straight-line code', which avoids program loops. However, this requires about 100 times the number of program steps, and therefore about 100 times the memory capacity [6]. With the processing times given in Table II, the processors can be used for many applications in the fields for which they were originally devel-

oped: telecommunications (especially in telephony) [9], audio and many kinds of speech-processing (such as speech coding, voice recognition and speaker identification).

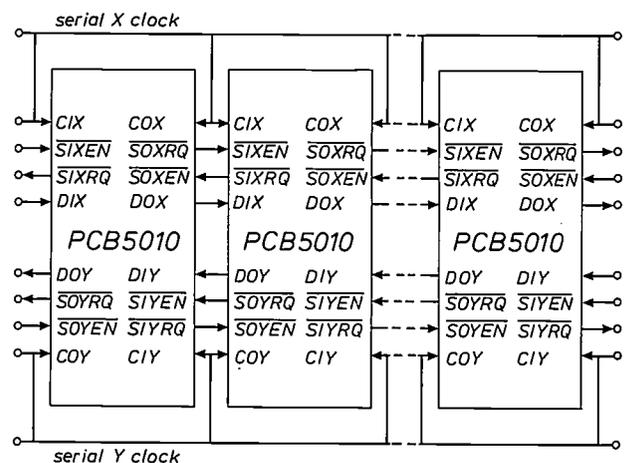


Fig. 16. Even greater versatility in processing can be obtained by combining several PCB 5010 chips. One method of combining the chips is illustrated here.

[9] K. Hellwig, K. Rinner, J. Schmid and P. Vary, Digitaler Signalprozessor für den Sprach- und Audiofrequenzbereich, PKI Tech. Mitt. No. 1, 57-64, 1986, Philips Kommunikations Industrie AG, Nuremberg, Germany.

Table II. Some benchmarks illustrating the signal-processing capacity of the PCB 5010/PCB 5011. The right-hand column gives the time required for performing a number of frequently occurring operations; in filtering operations this is the time required for calculating one sample of the output signal. Unless otherwise stated, all the quantities (signal samples, coefficients) have the standard word length of 16 bits.

Type of operation	Processing time (μ s)
Non-recursive filtering (per filter coefficient)	0.125
Non-recursive filtering (64 filter coefficients; including input and output)	9.250
Recursive filtering (2nd-order section)	0.625
Recursive filtering (2nd-order section)*	1.875
Recursive filtering (2nd-order section; including input and output)*	3.375
Complex multiplication	1
128-point FFT	927
128-point FFT (including window function and input and output)	1100
256-point FFT	2112
256-point FFT (including window function and input and output)	2300

* both signal samples and filter coefficients have double word length

Teamwork

The PCB 5010 is one of the results of the 'SIGMAPI' project. The team included staff from Philips Research Laboratories (Eindhoven), Valvo (Hamburg)

and TeKaDe (Nuremberg). Besides the author, F. J. A. van Wijk and F. P. J. M. Welten also shared the responsibility for the development of the chip architecture. They received considerable support from the system designers R. J. Sluijter, P. Vary and K. Hellwig. Others who contributed were A. Delaruelle, J. A. Huisken, J. Stoter, W. Gubbels, J. Schmid, K. Rinner and J. Wittek (in the design), and K. J. E. van Eerdewijk (in the testing).

Summary. Digital signal processors have gradually evolved away from the older computer concepts to become a separate class of large to very large digital integrated circuits. Modern versions have the 'Harvard architecture', which is characterized by separate arrangements for transfer and storage of data and control information. This also applies to the PCB 5010, developed primarily for applications in telecommunications, audio and speech-processing. The PCB 5010, fabricated in 1.5- μ m CMOS technology, contains 135 000 transistors on an area of 61 mm² and can execute eight million instructions per second. Each instruction takes the form of a 40-bit 'microcode' word and specifies a maximum of six different sub-operations that can be executed simultaneously. As a general rule the data words have a length of 16 bits, but for some intermediate results 40 bits are available and if required, computations can be carried out with greater precision. The PCB 5010 has three data memories (a 512 \times 16-bit ROM and two 128 \times 16-bit RAMs) and a program memory (1024 \times 16 bits, mostly in a ROM). Various items of supporting software and hardware are available to facilitate the application of the PCB 5010.

1938

THEN AND NOW

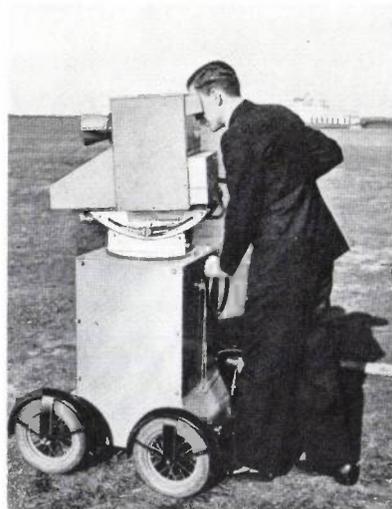
1988

Television cameras

In 1938 a transportable television camera was definitely something you did not see every day. Nor would many users ever have thought of taking one on holiday (lower photographs [*]). The pictures it took were only monochrome, of course. The zoom lens had not yet become a standard accessory, and there was still no simple way of recording the pictures. Since the number of picture lines had not been standardized, the camera shown offered the options of 405 or 567 picture lines rather than the 525 or 625 used today.

In 1988 many inventions in many fields have led to a very handy unit for the ordinary consumer. This is the 'camcorder', in which the camera and the recorder are combined. The very up-to-date VKR 6840 unit shown in the colour photograph only weighs 1.2 kg and measures 24 cm x 15 cm x 11 cm. The camera takes

colour pictures with sound, of course. Some of the essential functions like focusing, diaphragm control and setting the white balance have the option of manual or automatic operation. This camcorder has a zoom lens (zoom factor of 6) that can be operated by a motor or manually. The built-in video recorder (VHS-C system) gives an hour's 'filming' for each cassette. Recorded material can be inspected at any time with the built-in monochrome electronic viewfinder. The camcorder connects directly to a colour television receiver for playback in colour, or a cassette adapter can be used with any of the 130 million or more VHS recorders that have now been produced worldwide.



[*] From Philips Technical Review, January 1938.

Technological aspects of advanced telecommunications

G. Lorenz

1/4
5

The text below is an almost word-for-word account of the speech by Prof. G. Lorenz, member of the Group Management Committee of N.V. Philips' Gloeilampenfabrieken, at the European Conference 'Telecommunications: a European perspective', held on 18th and 19th June 1987 in the Kongresshalle in West Berlin. The speech is published by permission of the Corporate External Affairs Department of Philips International B.V. We have added the references and illustrations.

The speech gives an account of telecommunications today and looks ahead to future developments. We are sure it will interest our readers to learn how the management of our Company see these developments and how they think they should react to them from a European standpoint.

Modern telecommunications is going through a development phase that many people would describe as revolutionary. Throughout the world companies and research groups are working on these developments. Changes in communication are leading to the well-known convergence of communication and information. We can also see that the users are increasingly demanding global services and that they need low-cost, highly developed communication techniques that form the links between the terminals or between communication networks. Services intended for speech will continue to dominate for a long time to come, but even now the greatest growth area is to be found in non-speech services. The rapid increase in data communication will certainly continue. This year data communication within our company has increased by about 40%. Multimedia services intended for the combination of speech, data, text, pictures —

Prof. Dr G. Lorenz is a member of the Group Management Committee of N.V. Philips' Gloeilampenfabrieken.

either moving or stationary — are also being developed. In order that they can adapt to the needs of the market for non-speech services, the networks should have the option of flexible bandwidth, so that the services they offer can be adapted fairly easily by using different combinations. These developments in telecommunications, which will be of most interest to the business community, and will also be useful for the general public, particularly the private user, can only happen if the necessary technologies become a reality and consensus is reached on the standardized services and their coordinated introduction. We want more competition in Europe — so as to release dynamic forces. I think we would all endorse this, if the larger market, the unified European market, becomes a reality. I believe that the policy of the European Community is going in this very direction. If we only want to increase competition, without at the same time creating larger markets, we in Europe would be going in the wrong direction.

Telecommunications is based on three basic technologies: microelectronics, optical technology and software technology.

Everyone knows about the triumphal progress of microelectronics^[1]. Millions of electronic functions can be combined on a silicon chip (*fig. 1*): speeds are increased, dissipation reduced and the costs significantly cut. These were the developments that made

process this stored information at any place and at any time, and this is what we need for communication in dialogue form. Finally, the third option is an abundance of transmission capacity for all kinds of information. Nor is this last development at an end, and the basis of this unlimited transmission capacity is not only microelectronics, but optical-fibre technology as well.

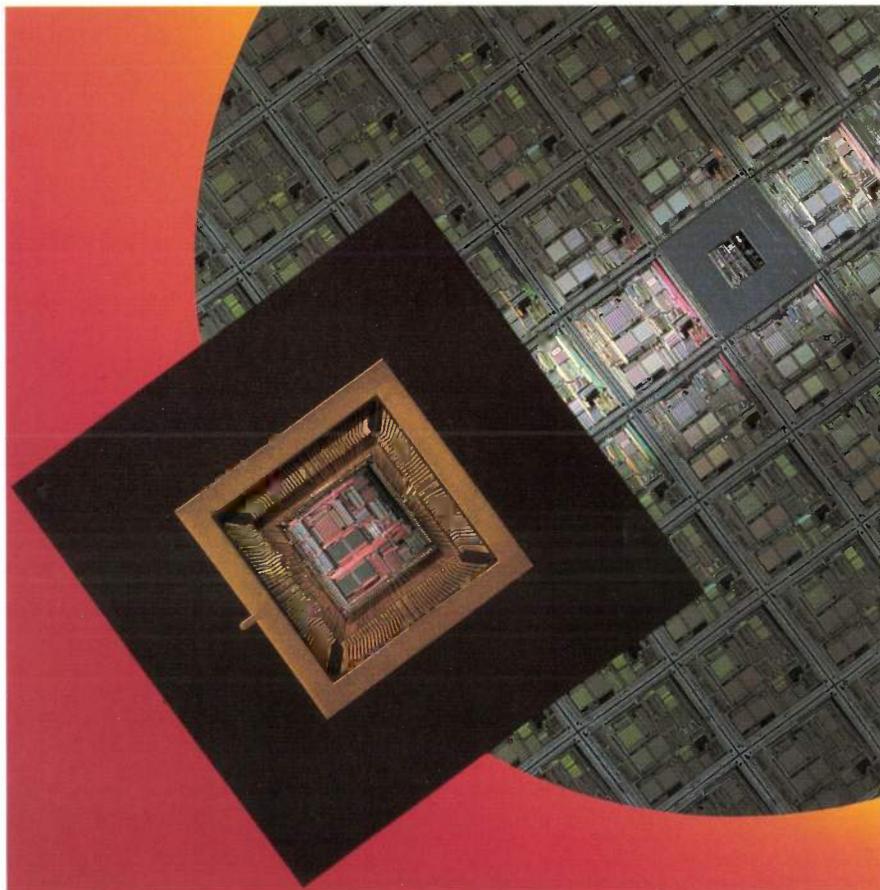


Fig. 1. Philips are producing increasing quantities of 'Very-Large-Scale Integration' circuits. These large and complex microelectronic circuits are true systems on a single chip. They require considerable technological capability and a great deal of 'know-how' relating to the applications of the chips.

ed. Philips Bedrijven B.V.
PHILIPS RESEARCH LAB
LIBRARY WY - 1
P.O. Box 80.000
5600 JA EINDHOVEN
THE NETHERLANDS

digitization^[2], integration and modern telecommunications into a reality. In the last 20 years the number of electronic functions on a chip has increased by a factor of 250 000 and the costs have fallen by a factor of 40 000. I believe that this is a unique occurrence in the history of industry. And these developments still continue, as we all know. Technological developments have given us another option: the amount of memory capacity. Today all kinds of information can be stored in large quantities at decentralized locations anywhere in the world. In principle, anyone can ac-

At present submicron technology is being developed throughout the world. We can assume that by the end of the eighties chips with many millions of transistor functions will be in mass production. By 1995 we — Philips, that is — expect to have structures with smallest dimensions of the order of 0.3 microns. This will be yet another step forward, which in turn

^[1] J. C. van Vessem, From transistor to IC: a long road?, Philips Tech. Rev. 42, 326-334, 1986.

^[2] See for example Philips Tech. Rev. 42, 101-144, 1985, Digital signal processing I, Background (special issue).

will offer a degree of integration of 100 million or more transistor functions. Finally, we can assume that the trend towards larger-scale integration will again be accompanied by a 10- to 100-fold reduction in the costs. These costs relate to the electronic functions.

The effect on the development of telecommunications is enormous: memory circuits and microprocessors are becoming more important, new processor families, in particular signal processors^[3], are being developed and produced everywhere, and parallel processing is the new buzz-word. Structures connected with Artificial Intelligence will gain in significance, as will ASICs (Application-Specific Integrated Circuits). Standard logic circuits will become less important, and so will customized circuits. We estimate that by 1995, 45% of all microelectronic circuits will be ASICs, and we expect that complete systems or subsystems will be integrated on a single chip. Systems-on-silicon is the in phrase. The condition for this development is not only that the technology should be available, but also that design and software know-how should be more effective, with the possibility of manufacturing prototypes quickly. As a result, the organizational structure of semiconductor companies will change. Until 1975 ICs were designed on the basis of geometric components, such as transistors, resistors, etc. Now design is based on structured components, such as registers or central processors. ASICs form the third stage, which is based on the concept of functional designs. Here is an example: work is in progress to enable a 'silicon compiler' for complex digital filters to produce the IC layout at the production centre from the filter equation alone. We are just starting to use this technology. It is in principle ready for use, the software is being improved and the development people are gradually learning to use it. ASICs will largely solve the problem of deciding between customized ICs — generally expensive because the quantities are small — and cheap standard circuits. As I have said, we have not yet reached the limits of silicon technology. We shall be there when we can integrate a billion (10^9) electronic functions. Then we shall have reached the physical limits of silicon technology. And this will happen in the foreseeable future.

The current digitization of telecommunications is essentially based on the microelectronic components available to us today. However, if we consider the future requirements for telecommunications, especially for broadband communication, we can see that we do not yet have the necessary technological conditions for making the circuits at economic cost. I have the impression that the innovation process in telecommunications is changing radically. Up till now microelectronics has been the motive force behind the innova-

tion process in telecommunications. But now we know that to an increasing extent telecommunications is the driving force behind the innovation process in microelectronics, and that is a fundamental change. It signifies a considerable responsibility for both researchers and development people in telecommunications.

The telecommunications of the future will require series-to-parallel converters, A/D and D/A converters, electro-optical converters, complex digital filters, cheap encoders and decoders, SLICs (Subscriber Line Interface Circuits), switches with fast clock rates. All these are components that cannot be used in their present form in computer technology. This is why I never tire of explaining the new product requirements to our IC producers, making clear to them that we need communication circuits, not just microcontrollers for memories and microprocessors. 140 megabits/second is the requirement of the future. Everyone working in this field knows what that means — clock rates of 1.2 GHz for time-multiplex switches. And for microelectronics this means gate propagation delays of 500 picoseconds. These requirements cannot yet be met with the current technology (CMOS). Bipolar technologies are faster, but today they are suitable only for small-scale integration, and gallium arsenide, as a new material being worked upon, will certainly be the basis of an important technology in the future, but this technology has the disadvantage that it is very expensive. Gallium arsenide is a compound, not a simple element like silicon, and it also does not seem fundamentally suited to complementary logic.

You will see that I have my doubts about gallium arsenide. True, its applications are still in their infancy, but I do not believe that within the next ten years it will compete with silicon on a broad basis: the development and innovation potential of silicon technology is still much too great. Gallium arsenide will certainly be used when very high speeds are required. Another alternative on which work is being carried out is the integration of bipolar transistors and CMOS processors. This provides circuits fabricated partly in bipolar technology where high speeds are required, and partly in CMOS technology where high speeds are unnecessary. These circuits are much better in terms of power consumption and permit the possibility of much larger-scale integration.

Future requirements for telecommunications will also depend upon the system technology chosen for broadband services. In this context I should mention the asynchronous time-multiplex process, i.e. fast packet switching. This technique has the advantage that it is in principle service-independent and gives the user the bandwidth he needs at a given time depending on the nature and quality of the service required. I be-

lieve that this development must be discussed in depth with the microelectronics manufacturers. In our opinion we must devote considerable attention to fast packet switching.

Now something about the optical technologies. Gallium arsenide will continue to be very important for lasers. We are working on further integration towards monolithic optoelectronic circuits based on

Now I should like to say a few words about fibre-optic communication^[41] (fig. 2). We have known for a long time that light is eminently suitable for the transmission of signals. The transmission of information by light was first suggested at the end of the last century. Only with the advent of the laser did it become possible to propagate light in a protected environment — the optical fibre. The optical fibre turns out

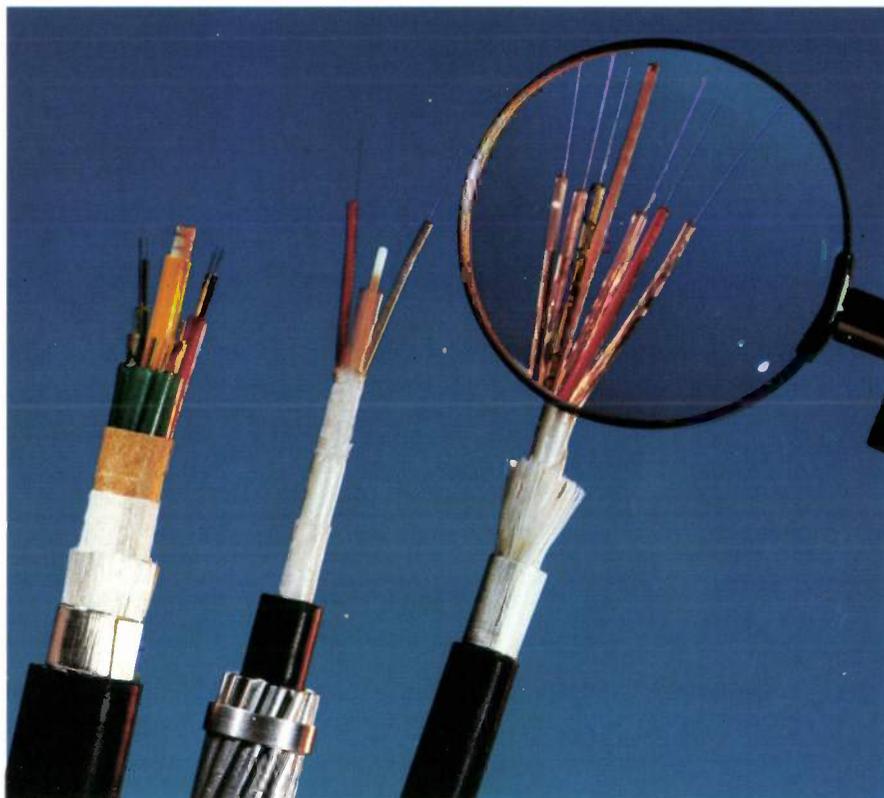


Fig. 2. Three optical-fibre cables. The left-hand cable contains six groups of ten optical fibres. The cable at the centre is a high-voltage cable intended for power distribution, with two optical-fibre cables at the centre for communication purposes. The right-hand cable contains six separate optical fibres. The optical fibres themselves are extremely thin, but each can carry many millions of bits per second.

substrates of new materials such as indium phosphide. At present experimental waveguide circuits — optical-waveguide circuits based on lithium niobate — are being developed. AT&T have recently presented studies of directly linked optical fibres - real 'photonics' therefore — but enormous R & D investment will be required before truly commercial 'photon' products such as switches, modulators, multiplexers and possibly even 'photon processors' and 'photon memories' come on to the market. We can already see that telecommunications will now bring forth photonics to stand beside electronics.

to be an ideal transport medium. Today fibre-optic cables can handle transmission rates of between 2 and 565 megabits/s. Distances of 35 kilometres can be bridged without repeaters. Fibre-optic cable is more economical than copper cable, with the result that the costs of information transmission fall. Because of its large bandwidth and insensitivity to external electro-

[3] See pp. 1-14 of this issue: J. L. van Meerbergen, Developments in integrated digital signal processors, and the PCB 5010.

[4] A. J. A. Nicia, An optical communication system with wavelength-division multiplexing and minimized insertion losses, I. System and coupling efficiency, Philips Tech. Rev. 42, 245-261, 1986.

magnetic fields, the optical fibre is more and more being used in local-area networks (LANs)^[6]. The LANs that we make and sell tend increasingly to be fibre-optic networks. We are actively working on the development of systems for 1.2 to 2.4 gigabits/s, and these will appear on the market in about two years. And there is a vast amount of research with coherent systems as the objective. We are also working on combining other optical functions such as modulation, polarization, equalization and directional coupling in integrated optics based on lithium niobate. In the future we will probably also include indium phosphide or gallium arsenide as the substrate in these projects. The introduction of the optical fibre enabled costs to be reduced, and this will continue. We are now working on optical fibres for wavelengths of between 2 and 3 microns and the same applies to the transmitters and detectors.

The optical fibre is an innovation. The application of optical fibres however is purely a process of substitution: something is replaced, it works a little more efficiently and is slightly less expensive. The real challenge for the optical fibre is to penetrate to the lower level of the transmission hierarchy. If we have an innovation, we must not let it become purely a matter of substitution. Such a substitution generally leads to lower sales and production and therefore fewer jobs. The important thing is that the creative application of an innovation should create products and markets — new services and equipment must come into being, new products and new services must be introduced in a creative way. The challenge for the optical-fibre technology is to penetrate the lower level right through to the private subscriber. We believe that each home will have an optical-fibre connection with a high transmission capacity^[6]. Today the cost factor still plays a fundamental role. But it is not only the cost factor, it is also the ideas factor, i.e. the answer to the question of what we are going to do with the optical-fibre option. Europe well understands the significance of the modern telecommunications infrastructure. The RACE programme is also aimed at the general introduction of optical-fibre networks from 1995 onwards. This programme needs the accompanying technology: the evolution of fabrication technologies, components, optical elements and finally process technologies, so that materials and equipment can become economically available. We must be ready for the transmission of moving pictures, probably of high-definition quality, via these networks.

I would like to say something here about optical memories, which are playing an ever-greater role. This is connected with the fact that the new telecommunications networks will incorporate more and more

intelligence, and new services will undoubtedly spring up. Databases and media for information storage will increase in importance and we think that the future need for broadband communication will be strongly influenced by the moving picture. If I may depart briefly from my main theme here: we talk a lot about moving pictures and other new services and we soon begin to wonder if there is a market for them, and if we need all of this. Of course we do not approach this problem with ordinary market research methods. Indeed, this applies to every basic innovation — by its very nature there never is a market for it at the start! As a rule, market research makes predictions for the future on the basis of existing markets. I can only say from experience that it is extremely difficult to predict how a basic innovation will fare in the market. I started in microelectronics in 1966. In that year we also wrote a scenario for 20 years ahead — 1986, and I was considered a hopeless optimist. But even this hopeless optimist's estimate was a factor of ten too low! This is why we must not give up just because there is no obvious market. As entrepreneurs we must not be faint-hearted. We must be bold enough to do innovative things. We must therefore tackle the new telecommunications a little more boldly and not concern ourselves too much with fundamental studies about future developments. Entrepreneurial initiative is stimulated by greater competition, of course, but it is stimulated even more by large markets. To put it another way: the unified European market is the real issue, and it is more important than greater competition. So I would like to give a higher priority to the European Community policy for the creation of a unified European market than to the call for more-competition in this market. However, I do admit that they are both related.

Let me now return to optical storage media (*fig. 3*) and mention the Compact Disc^[7], a real success story based on consumer electronics, not a professional application. We have already achieved submicron technology with our small silver disc. The little pits have structures smaller than one micron, so that millions of units of information can be stored in one square millimetre. We have also brought out CD-ROM. CD-ROM is used throughout the world for professional applications and has a memory capacity of 550 megabytes, which is equivalent to 1500 floppy disks or 200 000 standard A4 pages. And CD-ROM will be incorporated in the overall concept for office and production automation. We are bringing out a disc on which the user himself can enter data and as a next step we shall be bringing out a disc that can be rewritten many times. We have recently brought out CD-VIDEO, the integration of sound, text and vision.

An interactive Compact Disc (CD-I) is being prepared. We also have Digital Optical Recording (DOR), which enables large archiving systems^[8] to store 500 000 A4 pages or 25 000 pages in facsimile quality on an LP-size disc, corresponding to a resolution of

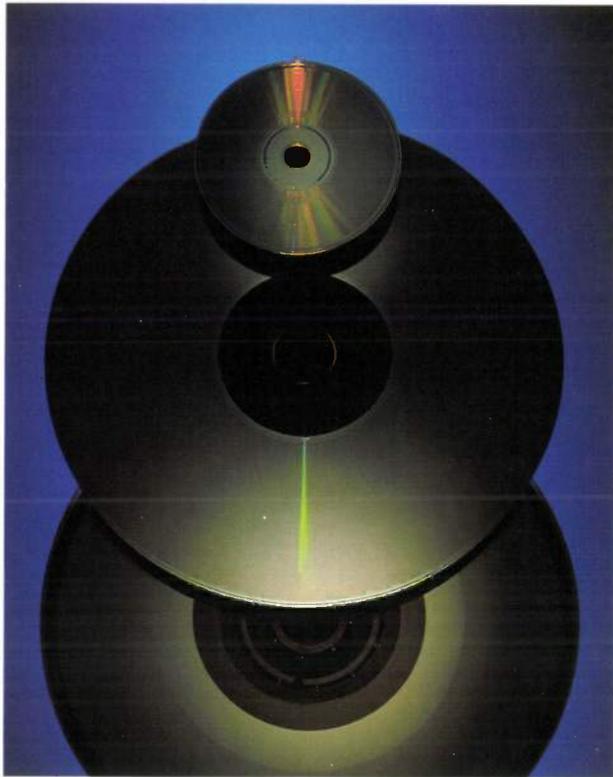


Fig. 3. This photograph shows various types of optical discs, all products of pioneering research work at Philips. An optical video disc is shown at the bottom of the photograph, and at the top a Compact Disc, now most familiar in the version used as an audio disc. At the centre there is an Optical Digital Data Disc, used for data storage.

4 million pixels per A4 page. Users of these systems work with many millions of A4 pages. They wish to enter information, read it, transfer it, and 'browse' quickly through it, and they want decentralized operation. And they want to be able to connect to the public network, too.

Now let me make a few more remarks about software technology. On the hardware side we have a quite astonishing costs trend that we do not find on the software side. This is one of the reasons why software is the problem today. We need better tools and methods for software development. We need a significant improvement in software productivity. Software productivity is increasing each year by about 10%: this is a not unusual increase, but for the rapid development

process in telecommunications it is not enough. Everyone knows that the software costs for the development of telecommunications systems now account for much more than 50% of the total costs. Work on this is under way at the research laboratories. We are examining methods based on formal specification procedures and on new design techniques — and in the context of ESPRIT and RACE^[9] programmes, too. We must intensify our efforts. In the research laboratories work is in progress on object-oriented programming and universal software, and finally application-oriented languages are being developed. Expert systems will play an important role in the highly-complicated field of test and diagnosis. It has to be said that the software technology is lagging behind. This is a very important point for the introduction of broadband communication as well as the technological aspects — the theme of this speech.

Another condition is the development of the market. We must achieve a united European market and we must also be bold enough to develop such markets, for example in the field of telecommunications. Another condition is the innovative strength of those offering services. They have a great responsibility. Innovations such as digitization, ISDN (Integrated Services Digital Network), microelectronics, optical fibres, optical technologies must not be seen as pure substitution processes. Modern telecommunications must be seen as a creative, innovative process. There must be a much closer cooperation between the user, the manufacturer and the provider of services, and in turn this must lead to new services being developed and offered. There will be the odd failure. There is no such thing as innovation without risk. We have tremendous opportunities, but there are also risks. Exploitation of the innovative potential is what counts, not merely substitution and rationalization. Compatibility is another condition, like standardization.

I would like to mention a final condition — a change in procurement policies. I think it is logical to say we want innovation, we must be bolder, but that also means we want larger markets, we want a unified

[6] J. R. Brandsma, PHILAN, a local-area network based on a fibre-optic ring, Philips Tech. Rev. 43, 10-21, 1986.

[7] J. van der Heijden, DIVAC — an experimental optical-fibre communications network, Philips Tech. Rev. 41, 253-259, 1983/84.

[8] M. G. Carasso, J. B. H. Peek and J. P. Sinjou, The Compact Disc Digital Audio system, Philips Tech. Rev. 40, 151-155, 1982.

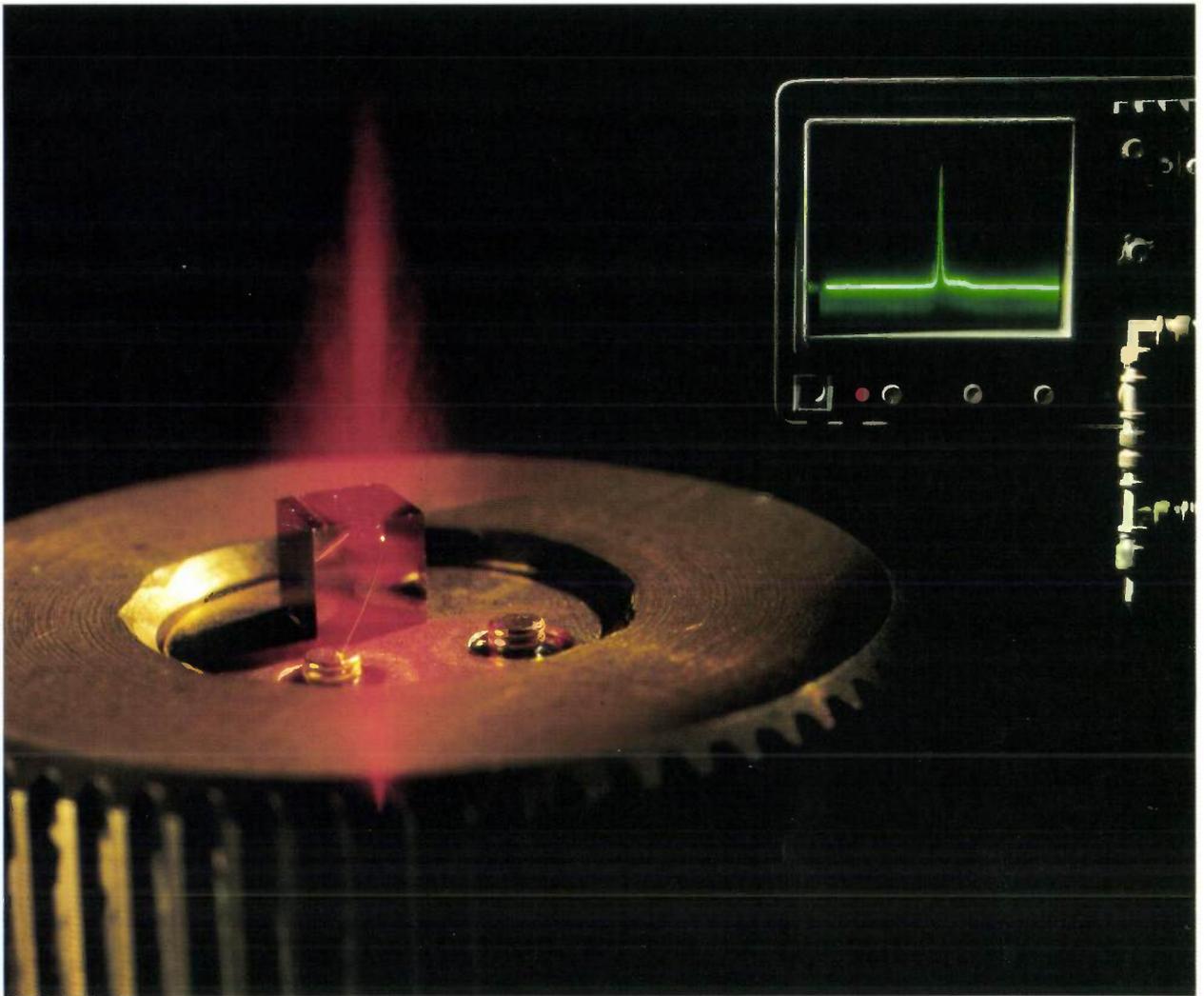
[9] L. Vriens and B. A. J. Jacobs, Digital optical recording with tellurium alloys, Philips Tech. Rev. 41, 313-324, 1983/84; J. A. de Vos, Megadoc, a modular system for electronic document handling, Philips Tech. Rev. 39, 329-343, 1980.

[9] ESPRIT: European Strategic Programme for Research and development in Information Technology.
RACE: R&D in Advanced Communications technologies in Europe.

European market. If we want a unified European market, we must be more consistent and demand a change in the procurement policies of the large European national governmental telecommunications departments. We welcome the European Community programme and recommendations for telecommunications.

To sum up: firstly, for us the basic technologies of microelectronics and optics represent unlimited resources for the future development of telecommunications. Electronic functions, memory capacity, transmission capacity are available in almost unlim-

ited quantities and inexpensively. The further development of technology will also create the conditions for broadband communication. Secondly, new technologies are being developed for integrated optics and photon-based components. Thirdly, there is a software bottleneck. New methods, increased productivity are required. Fourthly, the prerequisites for advanced telecommunications — apart from the technology, standardization, the development of new services — are a unified European system, a unified European market and changes in the national procurement policies.



Semiconductor laser for visible light

Lasers made of semiconductor materials are used as light sources in fibre-optic communication systems, laser printers and many kinds of optical recording systems, such as Compact Disc. In optical recording^[1] the wavelength of the laser light is of very great significance, since the maximum attainable information density on the optical disc is inversely proportional to the square of the wavelength. At present the laser wavelength usually has a value of about 800 nm, which is outside the visible-light range (400-700 nm). Recently Philips have made a semiconductor laser that operates at 650 nm. This laser was formed from a number of single-crystal layers (consisting of compounds of the chemical elements Al, Ga, In and P) with different compositions and doping. The layers were applied successively to a gallium-arsenide substrate by metal-organic vapour-phase epitaxy (MO-VPE)^[2]. The photo-

graph shows an operating laser without its protective casing. The shiny copper cube (2 mm × 2 mm × 2 mm) acts as a heat sink; the actual laser (0.3 mm × 0.3 mm × 0.1 mm) is attached to the top of the front surface, and can be identified by the connecting wire. This laser can provide light pulses at a peak power of 100 mW or more, and the limits of its performance have not yet been reached. Because of its high available power it can be used for writing data to optical discs as well as read-out. The oscilloscope picture at the upper right of the photograph shows that the spectrum of the emitted radiation has only one significant line, at 650 nm.

[1] See also G. E. Thomas, Future trends in optical recording, Philips Tech. Rev. 44, No. 2, April 1988.

[2] P. M. Frijlink, J. P. André and M. Erman, Metal-organic vapour-phase epitaxy of multilayer structures with III-V semiconductors, Philips Tech. Rev. 43, 118-132, 1987.

Improved adhesion of solid lubricating films with ion-beam mixing

K. Kobs, H. Dimigen, H. Hübsch and H. J. Tolle

The application of ion beams in semiconductor technology is widely known: the striking developments in the integrated-circuit industry would hardly have been possible without the progress in semiconductor doping achieved through ion implantation. In recent years there has also been increasing interest in other preparative methods, since ion bombardment is not only an excellent method of modifying the electrical properties of solids, but can also be used to modify their optical, chemical and mechanical properties. Furthermore, when thin-film deposition is combined with ion bombardment it is easy to produce novel surface alloys by ion-beam mixing, and problems due to poor adhesion between substrate and film can be avoided. At Philips GmbH Forschungslaboratorium Hamburg ion-beam mixing has been used to improve the adhesion of MoS_x lubricating films on steel to extend the useful life.

Introduction

A bombardment of a solid by a beam of energetic ions results in an implantation of ions, with displacements of atoms near the implanted ions. Ion implantation has been widely used for many years in the semiconductor technology for fabricating layers of n-type and p-type material [1]. In comparison with the more conventional method of doping by thermal diffusion it permits better control of the depth profile, with reduced lateral doping, and almost any element can be used as the doping material. These advantages more than outweigh the need for annealing and for equipment that is more complicated and expensive.

In recent years ion bombardment has been increasingly used for modifying metallic surfaces [2]. With the wide variety of ion beams and experimental conditions there are many ways of improving properties such as hardness, wear resistance and corrosion resistance. Ion bombardment can be performed at well-defined low temperatures, with hardly any effect on the manufacturing tolerances or the conditions for machining the surface.

Three methods for modifying a surface with ion beams are shown in *fig. 1*. In the first method there is an implantation very like that used for semiconductors, but at a much higher dose, of the order of 10^{17} cm⁻². This means that the concentration of the implanted ions may reach 10 at. % or more, leading to the formation of a surface alloy with a typical thickness of 0.1 to 0.3 μm, depending on the implantation conditions (the ion energy in particular) and the target material.

The other two methods in *fig. 1* make use of the 'mixing' capacity of an ion beam. In both, a deposition of a thin film is followed by ion bombardment to induce a kind of mixing of the atoms of the film and the substrate. This may produce some kind of surface mixture, or it may give a well-defined surface alloy (*fig. 1b*). The advantage over direct implantation (*fig. 1a*) is that the dose can be reduced by two orders of magnitude, because of secondary collisions between the atoms that have been 'knocked out' of their positions in the film or substrate. The experimental conditions can usually be adjusted, especially for somewhat thicker films, so that the mixing is restricted to the interface region of the film and the sub-

K. Kobs, Dr H. Dimigen, H. Hübsch and H. J. Tolle are with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

strate (fig. 1c). The adhesion of the film can be substantially improved in this way without affecting its surface properties.

Various examples of improved adhesion of metal films on substrates due to ion-beam mixing have been reported recently [3]. We have studied this method for improving the adhesion of solid lubricating films of MoS_x ($x = 1.6$ to 1.9) on steel [4]. In the last few years there has been increasing interest in the application of such films for reducing wear and friction. It has been shown that sputtered MoS_x films have excellent lubricating properties in inert gas or in vacuum [6]. With films obtained under optimized sputtering conditions the coefficient of friction can be as low as 0.02. Wider application is still limited, however, by the short 'sliding life', which depends on the adhesion of the film and the degree of cohesion between the crystallites of MoS_x [6].

In the ion-beam mixing of MoS_x films the effect on the film structure must also be taken into account, because of the high correlation with the friction. A film that is amorphous after bombardment no longer has lubricating characteristics, since the coefficient of friction will have increased to 0.4 [7]. It is therefore prudent just to modify the interface region. This requires careful optimization of the ion energy and dose for a given film thickness.

In this article we shall first present some of the results of our tribological investigations, which demon-

strate the beneficial effect of ion-beam mixing on the sliding life of MoS_x films. Then we shall discuss changes in the properties of the film due to the ion beams, and we shall conclude by considering the modification of the interface region.

Improving the sliding life

The equipment and the procedure for depositing MoS_x films on steel by r.f.-diode sputtering have been described in an earlier article in this journal [8]. The sputtering was performed in the non-reactive mode with an MoS_2 target and at an argon pressure of 2.6 Pa. The power density was 5 W/cm^2 , the residual gas pressure in the vacuum chamber was less than 6×10^{-4} Pa and the temperature was less than 60°C . Before starting a deposition the steel substrate was cleaned by ion etching to avoid any effects from surface contamination on the adhesion of the film.

The ion bombardment was carried out in a Varian-Extrion Model 200 implanter with a mass-separated ion beam (Fraunhofer Institut für Festkörpertechnologie, Munich, and the Fraunhofer Arbeitsgruppe für integrierte Schaltungen, Erlangen). The beam-current

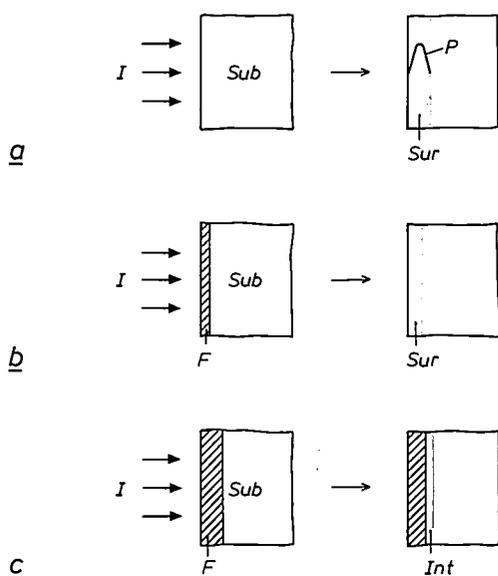


Fig. 1. Diagram showing three methods for ion-beam modification of solids [2]. a) Implantation of ions I into a substrate Sub , to produce a surface layer Sur of thickness 0.02 to $1 \mu\text{m}$ with a characteristic doping profile P . b) Mixing of the atoms of a thin deposited film F ($< 0.2 \mu\text{m}$) and the substrate to give a surface layer Sur of mixed composition. c) Formation of an interface layer Int of mixed composition between a film of thickness 0.05 to $1 \mu\text{m}$ and the substrate — this is the subject of this article.

- [1] See for example W. K. Hofker and J. Politiek, Ion implantation in semiconductors, Philips Tech. Rev. 39, 1-14, 1980; H. J. Ligthart and J. Politiek, An open 800-kV ion-implantation machine, Philips Tech. Rev. 43, 169-179, 1987.
- [2] See for example G. K. Wolf, Die Anwendung von Ionenstrahlen zur Veränderung von Metalloberflächen, Metalloberfläche 40, 101-105, 1986.
- [3] A. E. Berkowitz, R. E. Benenson, R. L. Fleischer, L. Wielunski and W. A. Lanford, Ion-beam-enhanced adhesion of Au films on Si and SiO_2 , Nucl. Instrum. & Methods Phys. Res. B7/8, 877-880, 1985; J. E. E. Baglin and G. J. Clark, Ion beam bonding of thin films, Nucl. Instrum. & Methods Phys. Res. B7/8, 881-885, 1985.
- [4] D. K. Sood, W. M. Skinner and J. S. Williams, Helium and electron beam induced enhancement in adhesion of Al, Au and Pt films on glass, Nucl. Instrum. & Methods Phys. Res. B7/8, 893-899, 1985; R. A. Kant, B. D. Sartwell, I. L. Singer and R. G. Vardiman, Adherent TiN films produced by ion beam enhanced deposition at room temperature, Nucl. Instrum. & Methods Phys. Res. B7/8, 915-919, 1985.
- [5] K. Kobs, H. Dimigen, H. Hübsch, H. J. Tolle, R. Leutenecker and H. Ryssel, Improved tribological properties of sputtered MoS_x films by ion beam mixing, Appl. Phys. Lett. 49, 496-498, 1986.
- [6] M. N. Gardos, Quality control of sputtered MoS_2 films, Lub. Eng. 32, 463-475, 1976; R. I. Christy and H. R. Ludwig, R. F. sputtered MoS_2 parameter effects on wear life, Thin Solid Films 64, 223-229, 1979; R. I. Christy, Sputtered MoS_2 lubricant coating improvements, Thin Solid Films 73, 299-307, 1980; H. Dimigen, H. Hübsch, P. Willich and K. Reichelt, Stoichiometry and friction properties of sputtered MoS_x layers, Thin Solid Films 129, 79-91, 1985.
- [7] T. Spalvins, Frictional and morphological properties of Au- MoS_2 films sputtered from a compact target, Thin Solid Films 118, 375-384, 1984.
- [8] T. Spalvins, Tribological properties of sputtered MoS_2 films in relation to film morphology, Thin Solid Films 73, 291-297, 1980.
- [9] H. Dimigen and H. Hübsch, Applying low-friction wear-resistant thin solid films by physical vapour deposition, Philips Tech. Rev. 41, 186-197, 1983/84.

density was $5 \mu\text{A}/\text{cm}^2$ and the maximum temperature was 150°C . The mixing experiments were performed with beams of singly charged nitrogen ions (N^+) or doubly charged argon ions (Ar^{2+}) at various energies (50 to 400 keV) and ion doses (0.3 to $5 \times 10^{16} \text{ cm}^{-2}$).

The friction and wear of the films were determined with a ball-on-disc tribometer with an oscillating steel ball [9]. The advantage of this method of testing is that it gives an unambiguous indication of the final failure of the film, which is signalled by a sudden increase of more than an order of magnitude in the coefficient of friction. The measurements were made in a dry nitrogen atmosphere with a relative humidity of less than 0.5%. Each film life was measured five times; the ball oscillated at a frequency of 7 Hz and the standard sliding load was 5.1 N.

The effect of the energy of the ion beams on the sliding life is shown in *fig. 2* for an MoS_x film $0.47 \mu\text{m}$ thick bombarded with nitrogen or argon ions at a dose of $1 \times 10^{16} \text{ cm}^{-2}$. No improvement is found at the lower energies; the mixing with nitrogen ions at 50 keV or argon ions at 100 keV in fact gives a slight reduction in the life. At high energies, however, there is a considerable improvement in the sliding life. The lives of nitrogen-bombarded films and argon-bombarded films are nearly identical, if we bear in mind that the nitrogen ions only require about half the energy of the argon ions.

The observed dependence of the sliding life on the energy of the ion beams indicates that the implanta-

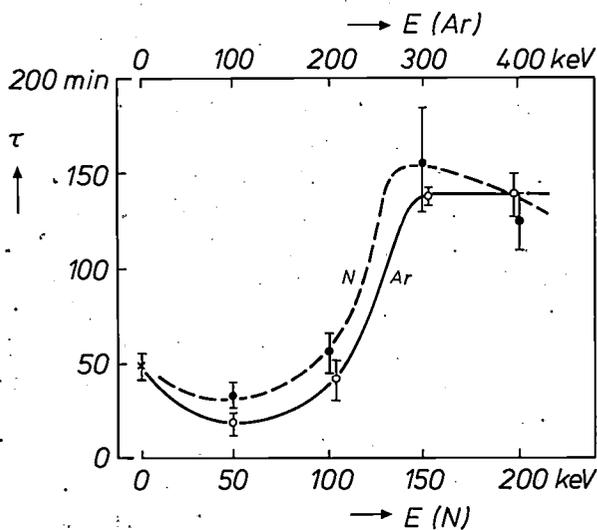


Fig. 2. Sliding life τ of $0.47\text{-}\mu\text{m}$ MoS_x films sputtered on steel, before ion-beam mixing (\times) and after mixing with nitrogen ions (\bullet) and argon ions (\circ), as a function of their energy E . The dose was $1 \times 10^{16} \text{ cm}^{-2}$, and the lives were measured at a sliding load of 5.1 N. Ion-beam mixing at high energies gives an appreciable improvement in the sliding life. The nitrogen ions require only about half the energy of the argon ions for the same sliding life.

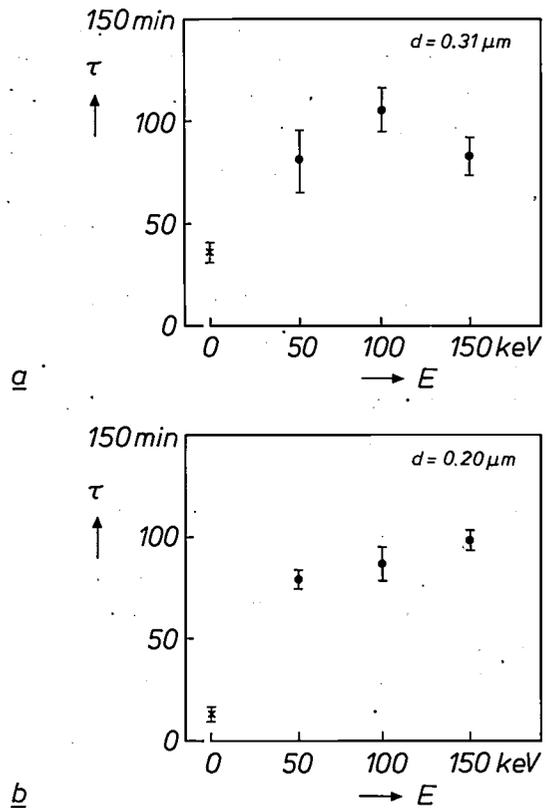


Fig. 3. Life τ at a sliding load of 5.1 N for MoS_x films of thickness $d = 0.31 \mu\text{m}$ and $0.20 \mu\text{m}$ sputtered on steel, before ion-beam mixing (\times) and after mixing with nitrogen ions (dose $1 \times 10^{16} \text{ cm}^{-2}$) as a function of their energy E (\bullet). Even after ion-beam mixing at a relatively low energy (50 keV) the sliding life is considerably improved — by a factor of more than two for the $0.31\text{-}\mu\text{m}$ film and by a factor of almost seven for the $0.20\text{-}\mu\text{m}$ film.

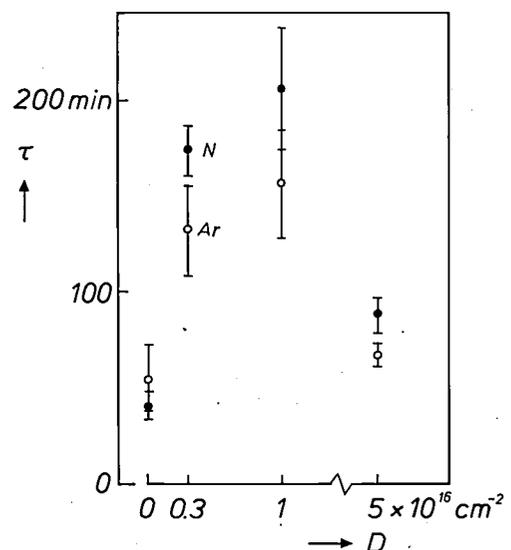


Fig. 4. Effect of the dose D on the life τ at a sliding load of 5.1 N for MoS_x films of thickness $0.43 \mu\text{m}$ sputtered on steel, after mixing with nitrogen ions at 150 keV (\bullet) and argon ions at 300 keV (\circ). In both cases the highest values are obtained at a dose of $1 \times 10^{16} \text{ cm}^{-2}$.

tion range of the ions is essential for the improvement in life due to interface mixing. This is in agreement with the results of life measurements for thinner films; see *fig. 3*. With a 0.31- μm film the life is found to more than double after mixing with nitrogen ions of energy only 50 keV. Under the same conditions the life of a 0.20- μm film increased by nearly seven times.

The dose dependence of the sliding life is shown in *fig. 4*. These results were obtained with 0.43- μm films bombarded with nitrogen ions at an energy of 150 keV and argon ions at an energy of 300 keV. The maximum life was observed at a dose of $1 \times 10^{16} \text{ cm}^{-2}$, but a very similar improvement was obtained at $3 \times 10^{16} \text{ cm}^{-2}$.

For practical applications it is very important to know whether the improvement in the sliding life is also found for higher loads. *Fig. 5* shows the effect of the load on the lives of three differently treated MoS_x films of thickness 0.29 μm deposited on steel. As the load was increased the life of the unbombarded film fell rapidly to only a few minutes. On the other hand, the sliding lives of films bombarded with nitrogen or argon ions were only reduced relatively slightly on increasing the load.

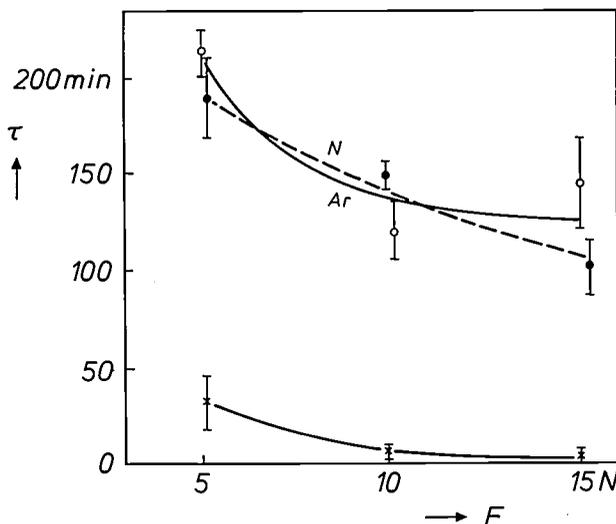


Fig. 5. The sliding life τ as a function of the applied load F for MoS_x films of thickness 0.29 μm sputtered on steel, before ion-beam mixing (\times) and after mixing with nitrogen ions at 100 keV (\bullet) and argon ions at 250 keV (\circ) at a dose of $1 \times 10^{16} \text{ cm}^{-2}$. At high loads the sliding life of the unbombarded film is reduced to a few minutes, whereas the bombarded films still have a long sliding life.

It should be emphasized that ion-beam mixing does not give any significant change in frictional behaviour during the sliding life. In all the tribological experiments it is found that the coefficient of friction is virtually unaffected by the ion energy and the dose [4].

Change in film properties

In seeking to explain the improvement in sliding life due to ion-beam mixing, we have used optical microscopy and profilometer measurements to study the wear tracks made by the oscillating steel ball. These investigations showed that there were considerable differences in the appearance of the bombarded and unbombarded films, which depended on the number of cycles of the oscillating ball. Even after a single cycle some of the material has been removed from the surface of the unbombarded MoS_x film. After ten cycles most of the wear track and the surrounding region have been damaged, indicating flaking. This has been confirmed by the profile measurements, which showed that the profile was rectangular.

These effects were not observed for MoS_x films bombarded with nitrogen or argon ions at high energies. For a 0.36- μm film bombarded with argon ions at an energy of 400 keV and a dose of $1 \times 10^{16} \text{ cm}^{-2}$, the first wear occurred after 1000 cycles with no flaking of the film. This indicates a considerable improvement in the adhesion of the film to the substrate. The wear track now has a typical ditch profile instead of a rectangular one. These results confirm observations of improved adhesion of sputtered MoS_x films that had been ion-bombarded at a very low film thickness before the sputtering had been completed [10].

Profilometer measurements of unbombarded and bombarded areas on the same sample reveal a significant reduction in the film thickness for the bombarded areas. This reduction increases with the energy of the incident ions; see *fig. 6*. The reduction in thickness is not due to sputtering effects, as has been demonstrated by electron-probe microanalysis. It is correlated with an increase in the film density of up to 40% and gives a greater cohesion between the MoS_x crystallites and an improvement in the effective film thickness.

A further indication of the effect of these structural changes is the increase in the reflectance of sputtered MoS_x films due to ion-beam mixing. For example, bombardment with argon ions at an energy of 400 keV gives an increase in the reflectance from 15 to 55%, probably because of the reorientation of the MoS_x platelets. It has been found that platelets with their basal planes perpendicular to the plane of the film reorient themselves parallel to the plane, giving an

[9] H. Dimigen, K. Kobs, R. Leutenecker, H. Ryssel and P. Eichinger, Wear resistance of nitrogen-implanted steels, *Mater. Sci. & Eng.* **69**, 181-190, 1985.

[10] J. Chevallier, S. Olesen, G. Sørensen and B. Gupta, Enhancement of sliding life of MoS_2 films deposited by combining sputtering and high-energy ion implantation, *Appl. Phys. Lett.* **48**, 876-877, 1986.

increased reflection^[11]. A partial recrystallization of the MoS_x film can also be induced by the ion bombardment, as has been demonstrated for sputtered WS₂ films^[12]. Investigations with scanning and transmission electron microscopes are in progress with the aim of clarifying the structural changes.

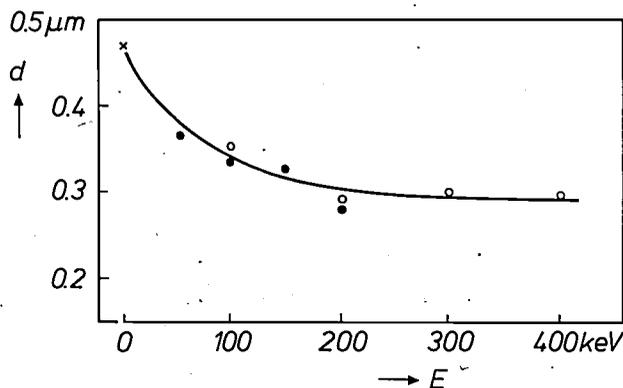


Fig. 6. Thickness d of MoS_x films sputtered on steel, before ion-beam mixing (x) and after mixing with nitrogen ions (•) and argon ions (o) as a function of their energy E at a dose of $1 \times 10^{16} \text{ cm}^{-2}$. A marked reduction in film thickness is observed for increasing ion energy.

Modification of the interface region

A useful method for studying the modification of the interface region between a film and its substrate is secondary-ion mass spectrometry (SIMS)^[13]. In this method an ion beam sputters ionized particles from the surface to be investigated. The ions sputtered successively are analysed by mass in a quadrupole field, and the variation in composition with depth can then be determined. Our SIMS experiments were performed by using a beam of oxygen or argon ions at an energy of 7 keV and a primary ion current of 0.5 μA , using a scanned area of 0.5 mm \times 0.5 mm. The resulting sputter rates were 1 $\mu\text{m/h}$ for the MoS_x films and 0.3 $\mu\text{m/h}$ for the steel substrates.

Fig. 7 shows the SIMS depth profiles of molybdenum, sulphur and iron for MoS_x films on steel after mixing with nitrogen ions at 50 and 200 keV. Mixing with ions at 50 keV does not give any broadening of the interface as compared with the unbombarded film. A markedly broadened interface of about 80 nm is detected after mixing at 200 keV, revealing a high correlation with the observed increase in the sliding life. The distance measured between the interface and the surface has become much smaller as a result of the reduction in the film thickness described previously.

Fig. 8 shows the SIMS profiles for the same type of film, but now after mixing with argon ions at 400 keV.

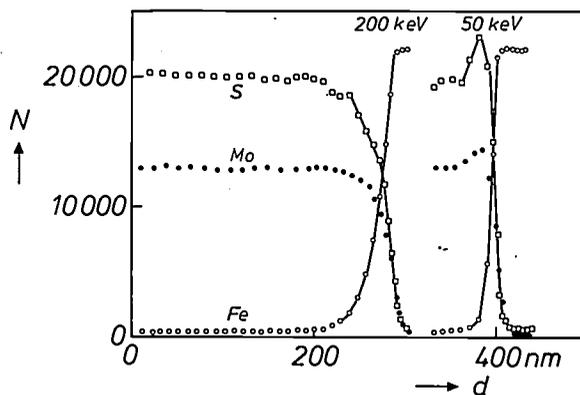


Fig. 7. SIMS depth profiles of molybdenum, sulphur and iron for MoS_x films of thickness 0.47 μm sputtered on steel, after mixing with nitrogen ions at 50 and 200 keV, at a dose of $1 \times 10^{16} \text{ cm}^{-2}$. The number of secondary-ion counts per second (N) is plotted against the depth d . After mixing at 50 keV some reduction in the thickness of the film is observed, but the intensity changes at the interface are fairly sharp. After mixing at 200 keV the reduction in thickness is much larger and the intensity changes are more gradual.

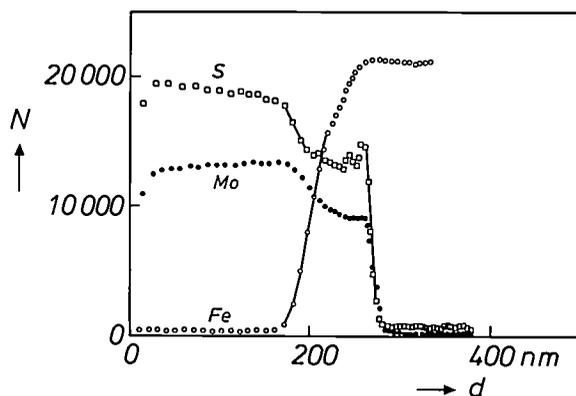


Fig. 8. SIMS depth profiles of molybdenum, sulphur and iron for an MoS_x film of thickness 0.47 μm sputtered on steel, after mixing with argon ions at 300 keV, at a dose of $1 \times 10^{16} \text{ cm}^{-2}$. The broadening of the interface region is more significant than in the samples of fig. 7.

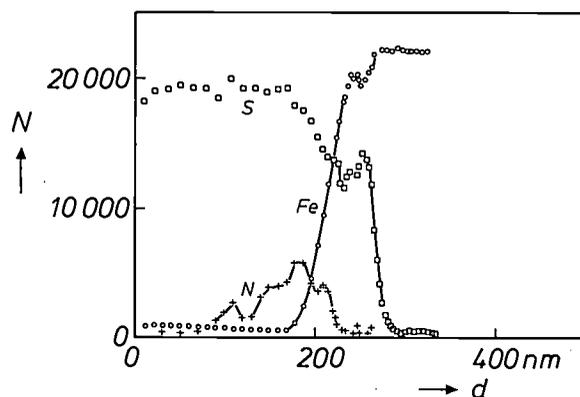


Fig. 9. SIMS depth profiles of sulphur, iron and nitrogen for an MoS_x film of thickness 0.43 μm sputtered on steel, after mixing with nitrogen ions at 150 keV, at a dose of $5 \times 10^{16} \text{ cm}^{-2}$. The intensity changes for sulphur and iron are almost the same as in fig. 8. The highest nitrogen intensity is observed at the interface.

The broadening of the interface is more significant than for the nitrogen-implanted films at the same dose, because of the greater mixing effect of the argon ions.

On increasing the dose of the nitrogen ions to $5 \times 10^{16} \text{ cm}^{-2}$ almost the same profiles are obtained as with the argon ions at a dose of $1 \times 10^{16} \text{ cm}^{-2}$; see *fig. 9*. The nitrogen depth profile was also determined for this sample. The maximum nitrogen concentration was found to be located at the interface. This indicates that the ion range corresponds well with the thickness of the MoS_x film.

In the work described here we cooperated closely with R. Leutenecker of the Fraunhofer Institut für Festkörpertechnologie, Munich, and with Prof. H. Ryssel of the Fraunhofer Arbeitsgruppe für integrierte Schaltungen, Erlangen.

The tribological tests were performed by W. Mohwinkel, and the electron-probe microanalyses were made by Dr P. Willich and D. Obertop.

Some of the work was supported by the Ministry of Research and Technology of the German Federal Republic, under Grants 03 T 0002 E6, 03 T 0002 A5 and 13 N 5353 /0.

[11] T. Spalvins, Morphological and frictional behavior of sputtered MoS_2 films, *Thin Solid Films* **96**, 17-24, 1982.

[12] M. Hirano and S. Miyake, Sliding life enhancement of a WS_2 sputtered film by ion beam mixing, *Appl. Phys. Lett.* **47**, 683-685, 1985.

[13] See for example K. Wittmaack, Successful operation of a scanning ion microscope with quadrupole mass filter, *Rev. Sci. Instrum.* **47**, 157-158, 1976;

H. W. Werner, Theoretical and experimental aspects of secondary ion mass spectrometry, *Vacuum* **24**, 493-504, 1974;

H. H. Brongersma, F. Meijer and H. W. Werner, Surface analysis, methods of studying the outer atomic layers of solids, *Philips Tech. Rev.* **34**, 357-369, 1974.

Summary. Ion-beam mixing of sputtered MoS_x films on steel can considerably lengthen their sliding life, especially at higher loads. The best results are obtained with films of thickness 0.3 to 0.5 μm bombarded with nitrogen ions at 150 to 200 keV or argon ions at 300 to 400 keV, with a dose of $1 \times 10^{16} \text{ cm}^{-2}$. The mixing has no adverse effect on the frictional behaviour during the sliding life. The improvement in the sliding life can be attributed to improved adhesion of the MoS_x films, as demonstrated by profile measurements and secondary-ion mass spectrometry. An ion beam can also introduce significant structural changes in a film, giving a noticeable increase in the film density and the reflectance.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

P. K. Bachmann, W. Hermann, H. Wehr & D. U. Wiechert	A	Stress in optical waveguides. 2: Fibres	Appl. Opt. 26	1175-1182	1987
J. M. Towner	S	Influence of conductor linewidth on short-circuit failure	Appl. Phys. Lett. 50	1581-1582	1987
J. M. De Blasi, M. Delfino, D. K. Sadana, K. N. Ritz & M. H. Norcott	S	Reduction of interfacial tunnel defects in silicon due to chemical vapor deposition of tungsten	Appl. Phys. Lett. 51	602-604	1987
P. H. L. Notten & A. A. J. M. Damen	E	The electrochemistry of InP in Br ₂ /HBr solutions and its relevance to etching behaviour	Appl. Surf. Sci. 28	331-344	1987
N. J. A. van Veen	E	XPS on impregnated cathodes: surface concentrations and thermal stability	Appl. Surf. Sci. 29	113-126	1987
H. Rau & W. Hermann	A	The nature of the reduced defect and the diffusion of GeO ₂ in germania-doped vitreous silica	Ber. Bunsenges. Phys. Chem. 91	833-840	1987
F. J. J. Blommaert* & J. A. J. Roufs* (*Inst. for Perception Res., Eindhoven)		Prediction of thresholds and latency on the basis of experimentally determined impulse responses	Biol. Cybern. 56	329-344	1987
E. W. Meijer & R. J. M. Zwiers	E	Photoinitiated cationic polymerization; a study to the polymerization and molecular mobility of poly-epoxide networks using phosphorescence spectroscopy	Crosslinked Epoxies, B. Sedláček & J. Kahovec (eds), Walter de Gruyter, Berlin	27-40	1987
M. Nikoonahad	N	Differential phase contrast acoustic microscopy using tilted transducers	Electron. Lett. 23	489-490	1987
P. J. Severin	E	Self-routing fibre-optic networks and switches using multitailed receiver/transmitter units	<i>ibid.</i>	651-652	1987
R. N. Bates & S. Feeney (Mullard, Stockport)	R	Novel varactor-tuned millimetre-wave gunn oscillator	<i>ibid.</i>	714-715	1987
D. H. Evans & R. N. Bates	R	Novel 93 GHz injection-locked impatt oscillator	<i>ibid.</i>	771-772	1987
J. G. Fijnvandraat	E	The effect of finite width in foil bearings: theory and experiment	Fluid Film Lubrication — Osborne Reynolds Centenary, D. Dowson <i>et al.</i> (eds), Elsevier Science, Amsterdam	539-542	1987
R. Clasen	A	Preparation of high-purity silica glasses by sintering of colloidal particles	Glastech. Ber. 60	125-132	1987
J. Opitz	A	Monosized powders for the preparation of high-purity silica glass	<i>ibid.</i>	133-139	1987

- P. J. A. Naus, E. C. Dijkmans, E. F. Stikvoort, A. J. McKnight*, D. J. Holland* (*Mullard, Southampton) & W. Bradinal (Valvo, Hamburg) *E* A CMOS stereo 16-bit D/A converter for digital audio IEEE. J. SC-22 390-395 1987
- O. Boser & H. Meehan (Magnavox, Mahwah, NJ) *N* Improved contacts in glass-encapsulated capacitors IEEE Trans. CHMT-10 103-106 1987
- J. E. Knowles *R* Anhyteresis in tapes: a Monte Carlo treatment IEEE Trans. MAG-23 201-203 1987
- B. J. Minnis *R* A printed circuit stub tuner for microwave integrated circuits IEEE Trans. MTT-35 346-349 1987
- M. L. G. Thoone, L. M. H. E. Driessen, C. A. C. M. Hermus & K. van der Valk *E* The car information and navigation system CARIN and the use of Compact Disc Interactive Int. Cong. and Exposition, Detroit, MI, Tech. Paper 870139 1-7 1987
- D. F. K. Hennings, R. Janssen (Tech. Univ. Hamburg-Harburg) & P. J. L. Reynen (Tech. Univ. Aachen) *A* Control of liquid-phase-enhanced discontinuous grain growth in barium titanate J. Am. Ceram. Soc. 70 23-27 1987
- H. J. Hagemann & H. M. O'Bryan (AT&T Bell Lab., Murray Hill, NJ) *A* Grain-boundary and surface segregation of Ba-Ti-O phases in rutile *ibid.* 274-278 1987
- R. Grössinger*, R. Krewenka*, H. R. Kirchmayr* (*Tech. Univ. Wien), P. Naastepad (Philips Metalware Factories, Eindhoven) & K. H. J. Buschow *E* Note on the coercivity in Nd-Fe-B magnets J. Less-Common Met. 134 L17-L21 1987
- K. H. J. Buschow & R. Grössinger (Tech. Univ. Wien) *E* Spontaneous volume magnetostriction in $R_2Fe_{14}B$ compounds J. Less-Common Met. 135 39-46 1987
- T. T. M. Palstra*, G. J. Nieuwenhuys*, R. F. M. Vlastuin*, J. van den Berg*, J. A. Mydosh* (*Univ. Leiden) & K. H. J. Buschow *E* Magnetic and electrical properties of several equi-atomic ternary U-compounds J. Magn. & Magn. Mater. 67 331-342 1987
- R. Clasen *A* Preparation and sintering of high-density green bodies to high-purity silica glasses J. Non-Cryst. Solids 89 335-344 1987
- C. A. M. Mulder & A. A. J. M. Damen *E* Raman analysis of the initial stages of the hydrolysis and polymerization of tetraethylorthosilicate J. Non-Cryst. Solids 93 169-178 1987
- M. J. Jongerius *E* Collisional broadening of the Na D lines by xenon in high-pressure sodium arcs J. Phys. B 20 3345-3365 1987
- R. M. Aarts *E* Algebraic expression for A-weighting response J. Sound & Vib. 115 372 1987
- B. Coulman & E. K. Broadbent *S* Interactions of Co thin films with Si substrates J. Vac. Sci. & Technol. A 5 1419-1420 1987
- G. N. A. van Veen, T. Baller, J. Dieleman & A. E. de Vries (FOM, Amsterdam) *E* Nanosecond ultraviolet laser induced etching of Si and Cu exposed to Cl_2 *ibid.* 1606-1607 1987
- C. T. Foxon, J. J. Harris, K. W. Barnham, D. E. Lacklison, J. Hewett & C. White *R* Two-dimensional electron gas with mobilities of $> 3 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ J. Vac. Sci. & Technol. B 5 778 1987
- M. L. Verheijke *E* Toepassing van neutronenactiveringanalyse, autoradiografie en radioactieve tracers in de siliciumtechnologie LAB/ABC 7 30-34 1986
- E. H. L. Aarts & H. M. Korst *E* Boltzmann machines and their applications Lecture Notes in Computer Science, Vol. 258, G. Goos & J. Hartmanis (eds), Springer, Berlin 34-50 1987
- H. Wehr & D. Wiechert *A* Refractive index and density of fluorine doped silica prepared by the PCVD process Mater. Res. Bull. 21 559-566 1986
- J. J. P. Bruines, R. P. M. van Hal, B. H. Koek, M. P. A. Vieggers & H. M. J. Boots *E* The transition between amorphous regrowth and explosive crystallization Mater. Res. Soc. Symp. Proc. 74 91-102 1987
- A. J. M. Houtsma* & H. J. G. M. Tholen* (*Inst. for Perception Res., Eindhoven) *E* A perceptual evaluation Music Perception 4 255-266 1987

- F. van Langevelde*, W. J. M. Lenglet*, R. M. W. Overwater*, R. D. Vis* (*Univ. Amsterdam), A. Huizing, M. P. A. Vieggers, C. P. G. M. Zegers & J. A. v.d. Heide (Univ. of Technol., Eindhoven) *E* X-ray focussing for synchrotron radiation microprobe analysis at the SRS, Daresbury (UK) Nucl. Instrum. & Methods Phys. Res. A257 436-442 1987
- G. Duggan, H. I. Ralph, P. Dawson, K. J. Moore, C. T. B. Foxon, R. J. Nicholas*, J. Singleton* & D. C. Rogers* (*Clarendon Lab., Oxford) *R* Transitions to confined states of the split-off band GaAs-(Al,Ga)As multiple-quantum-well heterostructures Phys. Rev. B 35 7784-7786 1987
- G. W. 't Hooft, W. A. J. A. van der Poel, L. W. Molenkamp & C. T. Foxon *E, R* Giant oscillator strength of free excitons in GaAs *ibid.* 8281-8284 1987
- G. Kimmel *N* High quality X-ray diffraction data using an adjustable divergence slit and thin samples Powder Diffraction 2 22-27 1987
- C. P. M. J. Baggen *E* Applications and problems of error correction coding with respect to storage channels Proc. 1st Int. Conf. on Industrial and applied mathematics, Paris-La Villette, 1987 183-196 1987
- R. N. J. Veldhuis & A. J. E. M. Janssen *E* Adaptive restoration of unknown samples in certain discrete-time signals; mathematical aspects *ibid.* 235-255 1987
- G. M. Martin, J. N. Patillon & C. Schiller *L* New material characterization requirements for III-V technology Proc. 6th Int. Summer School on Cryst. Growth, Edinburgh 1986 485-510 1987
- J. R. Mansell, D. Washington, A. G. Knapp, A. W. C. Hucks, R. A. M. Slater & D. A. Smith *R* The achievement of colour in the 12" channel-multiplier CRT Proc. Eurodisplay '87 77-80 1987
- K. G. Freeman, D. S. George & R. Pook *R* Circuits for the channel-multiplier flat, thin, colour CRT *ibid.* 216-218 1987
- C. M. Huizer *E* Optimized application of submicron CMOS for VLSI logic. A systems oriented view on design and technology Proc. IEEE Custom Integr. Circuits Conf., Portland, OR, 1987 391-395 1987
- G. Frens & C. A. M. Mulder *E* The effects of fluorine dope in fused silica Riv. Staz. Sper. Vetro No. 6 129-131 1986
- J. W. Orton, P. F. Fewster, J. P. Gowers, P. Dawson, K. J. Moore, P. J. Dobson, C. J. Curling, C. T. Foxon, K. Woodbridge, G. Duggan & H. I. Ralph *R* Measurement of 'material' parameters in multi-quantum-well structures Semicond. Sci. Technol. 2 597-606 1987
- J. W. M. Bergmans *E* Partial response techniques and robustness in data equalization Thesis, Eindhoven 1-194 1987
- J. W. C. de Vries *E* Temperature-dependent resistivity measurements on polycrystalline SiO₂-covered thin gold films Thin Solid Films 150 201-208 1987
- J. W. C. de Vries *E* Temperature-dependent resistivity measurements on polycrystalline SiO₂-covered thin nickel films *ibid.* 209-215 1987
- R. R. Manory (Lewis Res. Center, Cleveland, OH) & G. Kimmel *N* X-ray characterization of TiN_x films with CaF₂-type structure *ibid.* 277-282 1987
- E. K. Broadbent, A. E. Morgan, B. Coulman, I.-W. Huang & A. E. T. Kuiper *E, S* Characterization of titanium silicide films formed by composite sputtering and rapid thermal annealing Thin Solid Films 151 51-63 1987
- A. F. de Jong, H. Bender* & W. Coene* (*Univ. Antwerpen) *E* Actual comparison of experimental and simulated lattice images of the GaAs/AlAs interface Ultramicroscopy 21 373-377 1987
- A. F. de Jong, W. Coene* & D. van Dyck* (*Univ. Antwerpen) *E* On the phase-object function, used in dynamical electron diffraction calculations Ultramicroscopy 23 3-15 1987
- P. E. Wierenga & R. C. F. Schaake (P. D. Magnetics, Oosterhout) *E* The effect of mechanical and chemical surface properties on the friction of magnetic tapes Wear 119 29-50 1987



J. L. van Meerbergen, Developments in integrated digital signal processors, and the PCB 5010,
Philips Tech. Rev. 44, No. 1, 1-14, March 1988.

Digital signal processors have gradually evolved away from the older computer concepts to become a separate class of large to very large digital integrated circuits. Modern versions have the 'Harvard architecture', which is characterized by separate arrangements for transfer and storage of data and control information. This also applies to the PCB 5010, developed primarily for applications in telecommunications, audio and speech-processing. The PCB 5010, fabricated in 1.5- μm CMOS technology, contains 135 000 transistors on an area of 61 mm^2 and can execute eight million instructions per second. Each instruction takes the form of a 40-bit 'microcode' word and specifies a maximum of six different sub-operations that can be executed simultaneously. As a general rule the data words have a length of 16 bits, but for some intermediate results 40 bits are available and if required, computations can be carried out with greater precision. The PCB 5010 has three data memories (a 512 \times 16-bit ROM and two 128 \times 16-bit RAMs) and a program memory (1024 \times 16 bits, mostly in a ROM). Various items of supporting software and hardware are available to facilitate the application of the PCB 5010.

K. Kobs, H. Dimigen, H. Hübsch and H. J. Tolle, Improved adhesion of solid lubricating films with ion-beam mixing,
Philips Tech. Rev. 44, No. 1, 24-29, March 1988.

Ion-beam mixing of sputtered MoS_x films on steel can considerably lengthen their sliding life, especially at higher loads. The best results are obtained with films of thickness 0.3 to 0.5 μm bombarded with nitrogen ions at 150 to 200 keV or argon ions at 300 to 400 keV, with a dose of $1 \times 10^{16} \text{ cm}^{-2}$. The mixing has no adverse effect on the frictional behaviour during the sliding life. The improvement in the sliding life can be attributed to improved adhesion of the MoS_x films, as demonstrated by profile measurements and secondary-ion mass spectrometry. An ion beam can also introduce significant structural changes in a film, giving a noticeable increase in the film density and the reflectance.

G. Lorenz, Technological aspects of advanced telecommunications,
Philips Tech. Rev. 44, No. 1, 16-22, March 1988.

Word-for-word account of a speech at the European Conference 'Telecommunications: a European perspective' on 18th and 19th June 1987 in West Berlin. The speaker comes to the following conclusions. Firstly, for us the basic technologies of microelectronics and optics represent unlimited resources for the future development of telecommunications. Electronic functions, memory capacity, transmission capacity are available in almost unlimited quantities and inexpensively. The further development of technology will also create the conditions for broadband communication. Secondly, new technologies are being developed for integrated optics and photon-based components. Thirdly, there is a software bottleneck. New methods, increased productivity are required. Fourthly, the prerequisites for advanced telecommunications — apart from the technology, standardization, the development of new services — are a unified European system, a unified European market and changes in the national procurement policies.

I wish to subscribe to

PHILIPS TECHNICAL REVIEW

(date)

(signature)

Please tick the appropriate box

Regular subscription 80 guilders or U.S. \$ 35.00 per volume

The subscription includes postage and will start with Vol. 44, No. 1.
Please pay when you receive our invoice.

Student's subscription 32 guilders or U.S. \$ 14.00 per volume

Please send a copy of your student's card or other written proof that you are a student;
valid for two volumes.

Name

Initials

Title

Address

stamp
as
postcard

**Administration Department
Philips Technical Review**

**Philips Research Laboratories
Building WY 136**

P.O. Box 80 000

**5600 JA Eindhoven
The Netherlands**

Contents

256-kbit SRAM: an important step on the way to submicron IC technology	33
J. J. J. Bastiaens and W. C. H. Gubbels <i>Shrinking details and growing memories call for greatly modified technology</i>	
Then and now (1938-1988)	43
EXPERTISE: an expert system for infrared spectrum evaluation	44
T. Blaffert <i>The automation of methods of chemical analysis can save the chemist much time and trouble</i>	
Future trends in optical recording	51
G. E. Thomas <i>This overview shows that the main challenges will be increasing storage densities on the disc and speeding up reading and writing</i>	
Scientific publications	58

The 256-kbit SRAM: an important step on the way to submicron IC technology

J. J. J. Bastiaens and W. C. H. Gubbels

At Philips Research Laboratories development work has recently been completed on a Static Random-Access Memory (SRAM), a type of semiconductor memory, with a capacity of a quarter of a million bits. This 256-kbit SRAM has details as small as 1.2 μm and contains nearly two million transistors on a chip area of less than one square centimetre. The development of the 256-kbit SRAM constitutes an important step in the development of the Mega memory, which, with four times the capacity, will be a product of submicron IC technology.

Introduction

For some years now a race has been going on to build digital semiconductor memories of ever-increasing capacity. Philips have decided to take part in this race. Evidence of this decision is to be seen in the buildings of the submicron IC technology centre^[1], which have been erected at the Philips Research Laboratories site in Eindhoven; see *fig. 1*. The new buildings were officially opened in late 1986.

The objective is to develop and put into pilot production high-capacity semiconductor memories of the static-RAM type (Random-Access Memory). Cogent commercial considerations aside, an important motive for embarking on this work was that semiconductor memories of this type can act as the 'locomotive', to provide the motive power for submicron technology, i.e. IC technology with details smaller than 1 μm . These memories will play a major part in the learning process that must lead to an acceptable yield in the manufacture of various types of integrated circuits

with submicron details. Memory circuits also offer the advantage of enabling faults to be traced and localized very quickly: the address of a faulty cell gives a direct indication of the physical location of the fault. This greatly simplifies debugging procedures.

Successive generations of static RAMs — each corresponding to a quadrupling of the memory capacity — have appeared on the market at intervals of about three years. The total number of static RAMs sold has shown a faster percentage increase than that of dynamic RAMs. A cell of a static RAM (SRAM) consists of a flip-flop with four transistors, or two transistors and two resistors. Two additional transistors are required for connection purposes. A cell of a dynamic RAM (DRAM) is formed by a capacitor with one transistor. For the same level of technological sophistication, about four times as many DRAM cells as SRAM cells can be accommodated per unit chip area. In terms of packing density a 256-kbit SRAM is therefore comparable with a 1-Mbit DRAM.

Dr. J. J. J. Bastiaens and Ing. W. C. H. Gubbels are with Philips Research Laboratories, Eindhoven. The results described in this article relate not only to the work of the authors but also to the work of other colleagues at these Laboratories.

^[1] W. G. Gelling and F. Valster, The new centre for submicron IC technology, Philips Tech. Rev. 42, 266-273, 1986.

A drawback of a DRAM is that, because of charge leakage, the contents of the cells have to be periodically 'refreshed', which requires extra electronics on or off the chip. SRAMs require no extra electronics. Other advantages are that they are faster and draw less current. We have to remember here that both DRAMs and SRAMs are volatile memories, which means that the stored information is lost if the supply voltage fails.

The cells of the SRAMs dealt with in this article are made in full CMOS technology (CMOS stands for Complementary Metal-Oxide Semiconductor). This means that the actual cell consists of pairs of two types of field-effect transistor: NMOSTs and PMOSTs. In an NMOST the transistor operation is based on the formation of an n-type channel in p-type silicon, in a PMOST it is based on the formation of a p-type channel in n-type silicon. The cell consists of a combination of two logic inverters, each with an NMOST and a PMOST in series. In the steady state, i.e. when the cell does not change its logic state, one of the two transistors in each inverter is non-conducting. Current therefore only flows through the cell during the short time while the cell is changing from one logic state to the other. Consequently the current drawn by an SRAM cell in CMOS technology is relatively small.

The degree of 'sophistication' of the CMOS technology used can be characterized by the size of the smallest detail contained in the circuit. This is usually taken to be the length of the channel in the transistors. The SRAM considered in this article and recently

developed at Philips Research Laboratories has a channel length of $1.2\ \mu\text{m}$. The capacity of this memory is 256 kbit (i.e. $256 \times 1024 = 2^{18} = 262\ 144$ bits) and the chip area is $74\ \text{mm}^2$; see *fig. 2*.

Fig. 3 shows how the dimension d of the smallest details has evolved over the years [2]. There is always a 'phase lag' of about two years between production and development, and between development and research. Separate curves have therefore been drawn for production, development and research conditions; the curves have been extrapolated to 1990. The red 'point' in the figure indicates the packing density of the 256-kbit SRAM. The channel length in the SRAM of the next generation, the 'Mega memory', will be $0.7\ \mu\text{m}$. The capacity of this memory is 1 Mbit (i.e. $1024 \times 1024 = 2^{20} = 1\ 048\ 576$ bits). The Mega memory will be available earlier than *fig. 3* indicates: the first silicon wafers have gone through all the processing steps in the research stage successfully.

The special feature of the 256-kbit SRAM, over and above its fine detail and high capacity, is that it requires so little current. Its active current (the current required for reading and writing) is less than 15 mA and the stand-by current is less than $1\ \mu\text{A}$. Another feature is the low access time: 40 ns. The small active current and access time are due to the use of a circuit technique known as 'dynamic double-word-line' [3]. In our improved version of this technique less current is required both for reading from a cell and for writing information into it.

As a provisional result of the development work on the 256-kbit SRAM various types of 64-kbit SRAM,



Photo: Flying Camera

Fig. 1. The buildings of the centre for submicron IC technology at the Philips Research Laboratories site in Eindhoven were officially opened on 2nd December 1986 by H. R. H. Prince Claus of the Netherlands, by symbolically 'baking the first chip'.

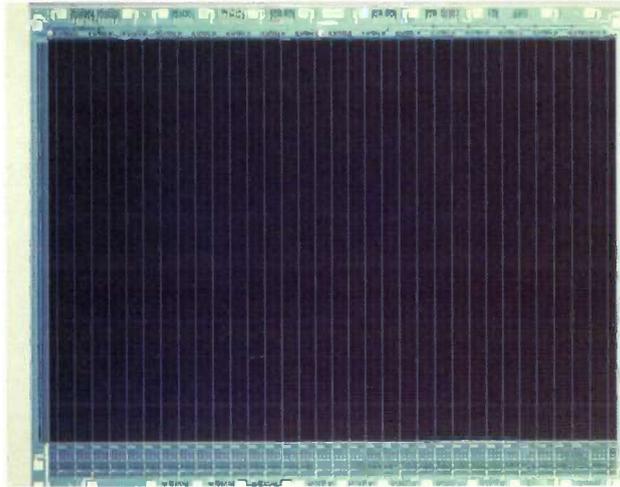


Fig. 2. Photograph of the 256-kbit SRAM developed at Philips Research Laboratories. Around the actual memory matrix with its $256 \times 1024 = 262\,144$ cells are the electronic circuits for addressing the cells and writing and reading the contents.

made with the same technology and circuit techniques, will shortly go into pilot production. Since they have a smaller chip area than the 256-kbit SRAM, these memories will have a higher production yield. With a quarter of the capacity, they will require less current and their access time will be slightly lower. The current is very much less than that of the 64-kbit SRAMs now on the market, which are not made in full CMOS technology but are built up from cells of two transistors and two resistors.

In the rest of the article we shall take a closer look at the CMOS technology on which the 256-kbit SRAM is based. The principle on which information is read from and written into a cell will then be explained. Finally, some of the difficulties encountered in scaling down the dimensions of MOS transistors to $1\ \mu\text{m}$ or less will be discussed — and ways of overcoming them will be indicated.

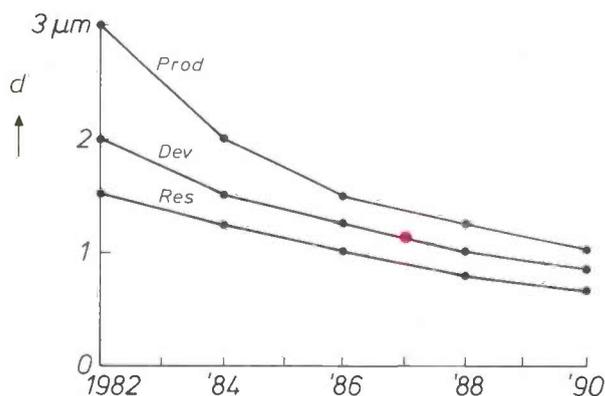


Fig. 3. The dimension d of the smallest details in integrated circuits through the years^[2], for research (Res), development (Dev) and production (Prod), extrapolated to 1990. The red 'point' refers to the 256-kbit SRAM.

The 256-kbit SRAM

The cell

A cell of the new 256-kbit SRAM consists of six transistors: two NMOSTs and two PMOSTs for the actual cell and two NMOSTs for the connections to the bit lines. The bit lines are used for addressing the columns of the memory, which is formed as a matrix in the conventional way. The rows are addressed via 'word lines'. The two logic inverters that constitute the actual cell form a flip-flop with two stable states: '1' and '0' (see fig. 4a and b). T_1 and T_2 are NMOSTs, and T_3 and T_4 are PMOSTs. Transistors T_3 and T_4 act as loads for T_1 and T_2 respectively.

When $V_{dd} = 5\ \text{V}$ and $V_{ss} = 0\ \text{V}$ the logic state '1' could typically be the state where the voltage at the point C_5 is high and the voltage at C_6 is low. Since T_1 is an NMOST, it conducts when the voltage on the gate is high, and is then 'off' in state '1'. Since T_3 is a PMOST, it conducts when the gate voltage is low, and it is therefore 'on'. For T_2 and T_4 the opposite situation applies. In the logic state '0' the voltage of point C_5 is low and that of C_6 high. In the steady state neither branch of the flip-flop conducts. The only currents are leakage currents.

Fig. 4c shows a micrograph, made with a Philips scanning electron microscope, of a cell of the 256-kbit SRAM in the stage of fabrication in which the first layer of aluminium lines is applied. These include the conductor lines for supplying power to the cell (they are horizontal in the figure). The second layer of aluminium lines (vertical) contains the bit lines, which are not visible in the micrograph. The cell is rectangular with dimensions of $8\ \mu\text{m} \times 25\ \mu\text{m}$. Fig. 4d shows the outline of the masks required in reaching the fabrication stage in fig. 4c. The NMOSTs T_1 and T_2 are insulated from the PMOSTs T_3 and T_4 by a relatively thick silicon-dioxide layer, referred to as field oxide (FOX) in fig. 4b. The field oxide is applied by the LOCOS technique (local oxidation of silicon)^[4], which we shall deal with in more detail. The rectangular shape of the cell permits efficient utilization of the surface, especially since the row-selection line RS (to be mentioned later) can be located over the field oxide between the NMOSTs and PMOSTs. So that

[2] J. Burnett, Clean rooms for ULSI manufacturing: class 1 practice, *Solid State Technol.* **28**, No. 9 (September), 121-123, 1985.

[3] T. Sakurai *et al.*, A low power 46 ns 256kbit CMOS static RAM with dynamic double word line, *IEEE J.* **SC-19**, 578-585, 1984.

[4] J. A. Appels, H. Kalter and E. Kooi, Some problems of MOS technology, *Philips Tech. Rev.* **31**, 225-236, 1970; B. B. M. Brandt, W. Steinmaier and A. J. Strachan, LOCOS, a new technology for complementary MOS circuits, *Philips Tech. Rev.* **34**, 19-23, 1974; S. M. Sze (ed.), *VLSI technology*, McGraw-Hill, New York 1983.

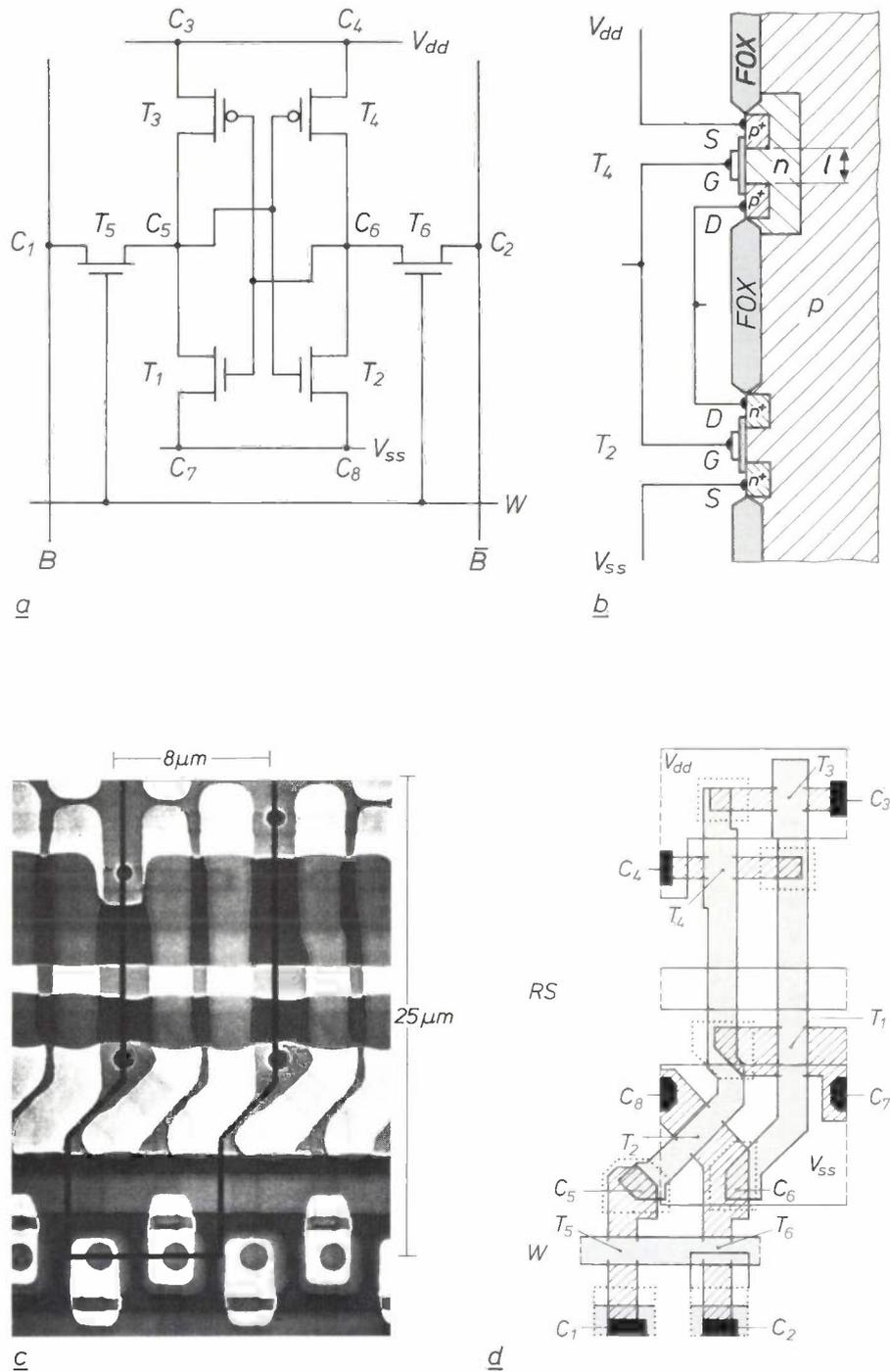


Fig. 4. A cell of the 256-kbit SRAM. *a*) The principle of the electronic circuit: two logic inverters form a flip-flop. B bit line, \bar{B} inverse bit line. W word line. T_1 to T_6 transistors. (T_3 and T_4 are PMOSTs, the others are NMOSTs.) V_{dd} and V_{ss} high and low supply voltages. C_1 to C_8 connection points, which can be found in (*d*). *b*) Schematic cross-section through the silicon wafer at the location of the inverter formed by T_2 and T_4 . FOX field oxide. n silicon with a surplus of free electrons, p , p^+ silicon with a surplus of free electrons, p , p^+ silicon with a surplus of holes respectively. l length of a transistor channel. S source. G gate. *c*) Micrograph made with a Philips scanning electron microscope. The cell outlined is in the fabrication stage in which the first layer of aluminium lines (here shown horizontal) for the power supply is produced. The vertical lines for the gates are of polycrystalline silicon. *d*) The masks required for reaching the fabrication stage in (*c*). The masks for the electrode diffusion regions are shown hatched, those for the polysilicon lines are grey and those for the first layer of aluminium lines are white. The black areas are aluminium contacts with diffusion regions. The areas surrounded by a dotted line are polycrystalline silicon contacts with diffusion regions. Most of the contacts are also shown in (*a*). Some of the masks end at dashed lines because they run through to neighbouring cells. RS row-selection line. The (vertical) bit lines are not shown.

the layout of the cell on the silicon surface can be compared with the actual electronic circuit, fig. 4d also shows the connection points from fig. 4a.

A cell consisting of two transistors and two resistors to replace T_3 and T_4 in fig. 4a is an alternative to the SRAM cell with two NMOSTs and two PMOSTs. The advantage of this widely used alternative is that it saves some space on the chip, since the resistors, fabricated in polycrystalline silicon, are in a separate layer above the two cell transistors and the two connecting transistors. A serious disadvantage, however, is that in the steady state there is always a current in one branch of the flip-flop, in spite of the high resistance of these resistors. The dissipation of such an SRAM with resistive loads is therefore always higher than that of an SRAM made in full CMOS technology.

Another disadvantage of an SRAM cell with resistors is that there is a higher probability of 'soft errors', which occur when α -particles strike a transistor channel. These particles may be of cosmic origin, or may come from the packaging or from impurities in the aluminium connection lines. Such a cell may also change state more easily when holes or electrons are injected into the channel from the substrate or when there is noise in the supply voltage of an individual cell. These effects are even more troublesome in a cell of a dynamic RAM.

Another argument in favour of an SRAM cell with two NMOSTs and two PMOSTs, i.e. in full CMOS technology, should not be discounted: full CMOS technology is much more compatible with the technology used for other ULSI logic circuits (ULSI stands for ultra-large-scale integration) that will result from the SRAM projects.

Writing to and reading from a cell

The integrated circuit contains a row-address decoder and a column-address decoder to find the correct cell in the memory matrix. The row-address decoder translates the row address and at the right moment it makes the word line W of the addressed row 'high'. (The supply voltage of 5 V is applied to the word line.) The transistors T_5 and T_6 in all the cells of the row then 'switch on'. The column-address decoder translates the column address and makes a connection to the bit line B and the inverse bit line \bar{B} of all the memory cells in the column addressed. When information is to be written into a cell, B becomes high and \bar{B} becomes low for the logic state '1'. For state '0', B becomes low and \bar{B} high.

When information is to be read from a cell, B and \bar{B} of the appropriate column are both made high. If the cell is in state '1', then T_1 is off and T_2 on. If the word

line of the addressed cell becomes high, a current starts to flow via \bar{B} and T_6 to the connection with the voltage level V_{ss} . The level of \bar{B} then becomes lower and that of B remains high. Similarly, if the cell is in state '0', \bar{B} remains high and B becomes lower. One of the two results is supplied via a differential amplifier and a buffer to the output of the integrated circuit.

The 256-kbit SRAM is arranged as a memory with a 'width' of 8 bits. This means that there are in reality not 2^{18} addresses but $2^{18}/8 = 32\,768$, each corresponding to eight cells in which 8 bits (1 byte) can simultaneously be written in parallel, or from which 8 bits can be read. Two spare rows and 32 spare columns are added to the memory matrix, which can be connected in if any rows or columns contain faulty cells. Rows or columns can be connected in or disconnected by interrupting conductor lines with a focused laser beam. The production yield can be increased in this way.

The memory matrix proper consists of 256 rows and 1024 columns. Because of the capacitance between a word line and the substrate, using 1024 cells per word line would give too large a delay in addressing and would therefore make the access time too long. A large number of cells per word line also means high dissipation, since the cells simultaneously addressed by a high word line start to take current via T_1 or T_2 and a bit line. The memory matrix is therefore divided into 32 sections of 32 columns and 256 rows, as shown in fig. 5. The number of cells to be addressed by one word line is thus reduced to 32.

With the memory matrix divided into sections in this way, the addressing of eight bits or one byte becomes a more complicated process than for addressing one bit. The circuit contains 256 row-selection lines RS (horizontal in fig. 5), which can address the required row in a section. There are also 32 (vertical) section-selection lines SS , for addressing the required section. At the point where a high row-selection line crosses a high section-selection line a word line W is made high by a logic circuit LC , so that 32 cells in a row of that section are addressed. The column-address decoder selects 8 of these 32 cells. The 8 selected cells are read or written in parallel. The procedure described is known as the 'double-word-line technique'.

The dissipation of the memory is kept low by energizing the appropriate word line for only 50 ns in each write or read cycle. When a cell is read this period of 50 ns does not start until an address changes. This is called the 'dynamic double-word-line technique' [3]. A special feature of the Philips 256-kbit SRAM is that one of its options is a circuit that can save current during the *writing* process as well. The saving is achieved by not starting the 50-ns period

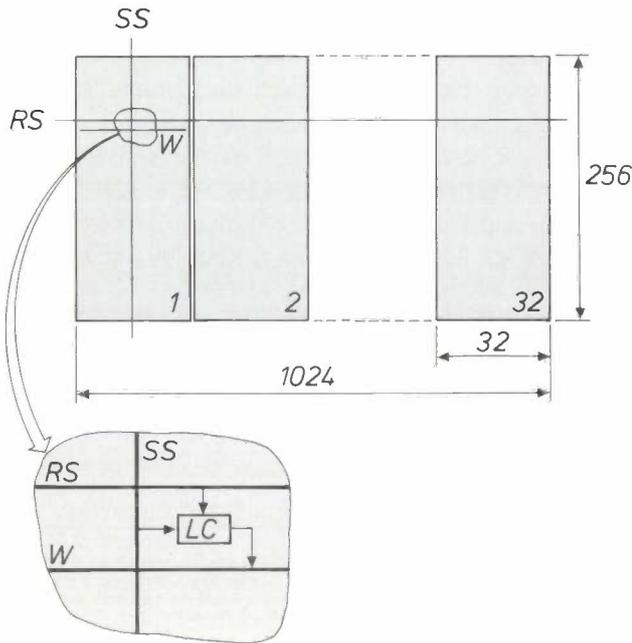


Fig. 5. The matrix of the 256-kbit SRAM consists of 256 rows and 1024 columns. The matrix is divided into 32 sections of 32 columns. *RS* row-selection line. *SS* section-selection line. At each point where a row-selection line crosses a section-selection line there is a logic circuit *LC* that determines whether the word line *W* at that intersection should be made 'high' or not. There are therefore $32 \times 256 = 8192$ of these logic circuits.

until a new byte is presented at the input. The circuit can be set up during manufacture by interrupting a line with a laser beam as mentioned above.

Problems in scaling down details

General

In the production of very large semiconductor memories a batch of silicon wafers is subjected to more than 200 consecutive steps in the process. The total time required is a few months. The processing includes 14 or more 'masking steps', in which a photosensitive resist layer on the wafer is exposed. If a 'positive' photoresist is used, the *exposed* areas are subjected to a further processing step, such as etching or ion implantation. If a 'negative' photoresist is used, it is the *unexposed* areas that are subjected to the next step in the process.

As chip areas tend to grow larger and the details in the circuit tend to become smaller, it is very important to control the amount of dust and particles during the entire process. Fig. 6 shows how the permissible defect density per masking step has diminished in recent years^[2]. The defect density plotted here is the minimum required to achieve a 10% yield of working circuits per wafer. The figure of 0.05 faults per cm^2 per masking step in 1990 means that an IC plant now

being set up will in future have to allow no more than 140 particles of $0.2 \mu\text{m}$ diameter or larger to land on a 6-inch silicon wafer. This is an extremely difficult specification for the cleanliness of the air, liquids and gases used during manufacture. The process operators will also have to wear specially designed protective clothing. Mechanization will be used to reduce human contact with the wafers to the minimum during the fabrication process.

Fig. 7 shows a micrograph made with a Philips scanning electron microscope of a cross-section of part of a cell of the 256-kbit SRAM. Three of the four interconnection layers can be seen. The first interconnection layer, n^+ and p^+ diffusions for the source and drain electrodes of the transistors, is not visible in the micrograph, but is shown schematically in fig. 4b. The second interconnection layer, of polycrystalline silicon and designated *PS*, contains the gate electrodes and their connections; this layer can be seen clearly in fig. 7. (Heavily doped polycrystalline silicon is a good conductor.) The third interconnection layer, of aluminium and designated *Al1*, connects the cells to the voltages V_{dd} and V_{ss} ; see fig. 4. The fourth layer, also of aluminium and designated *Al2*, contains the bit lines. The figure also shows a contact between layer *Al1* and the first layer produced by diffusion. This contact is used for applying a voltage to a transistor electrode. It is produced by etching a hole in the insulating silicon dioxide (light grey in the picture).

It is obvious that the relative alignment of the different layers must be highly accurate. Conductor lines or contact holes connected to different voltages must never be short-circuited. This means that the exposure for each masking step for a wiring pattern must be performed extremely accurately. The total permitted deviation is equal, with a probability of 99.7%, to three times the standard deviation σ_t of the sum of a number of random errors: the random variations of the width of a line, of an adjacent contact area and of the positioning of the two masks. This is illustrated in fig. 8, where the value of $3\sigma_t$ is plotted as a function of detail size. The graph is based on Philips design rules, which presuppose the use of a wafer stepper for the exposure of the wafers rather than a projector that illuminates the entire wafer in a single exposure.

For the future Mega memory — with details of $0.7 \mu\text{m}$ — a value as low as $0.35 \mu\text{m}$ is specified for $3\sigma_t$. This calls not only for accurate linewidth control of the processes but also for an extremely accurate wafer stepper^[5]. It seems likely that optical techniques will give satisfactory results for submicron technology until well into the nineties. After then it will be necessary to change over to techniques using X-rays, electrons or ions.

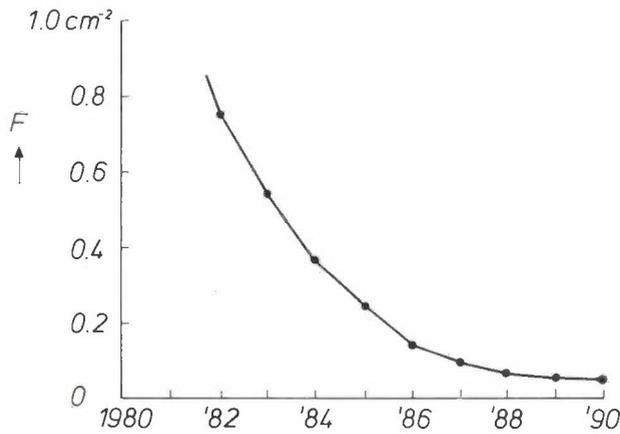


Fig. 6. The permissible defect density F per masking step through the years^[2].

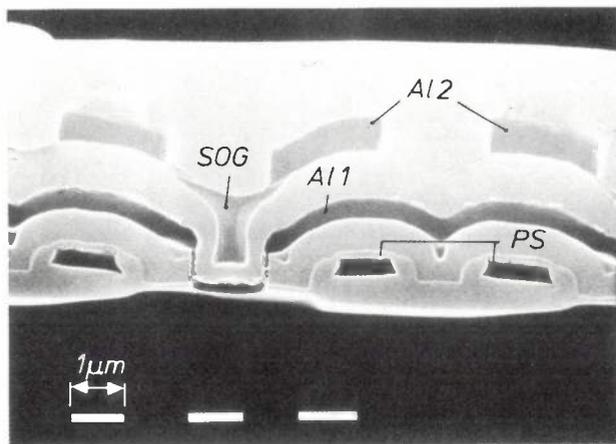


Fig. 7. Micrograph made with a Philips scanning electron microscope of a cross-section of a cell of the 256-kbit SRAM. Three of the four interconnection layers can be seen. *PS* polycrystalline silicon of the second layer. *Al1* aluminium of the third layer. *Al2* aluminium of the fourth layer. The light-grey areas represent silicon dioxide. *SOG* silicon dioxide applied to the spinning wafer ('spin-on glass') to improve the uniformity of the wafer. This silicon dioxide fills the opening left after a contact hole has been made.

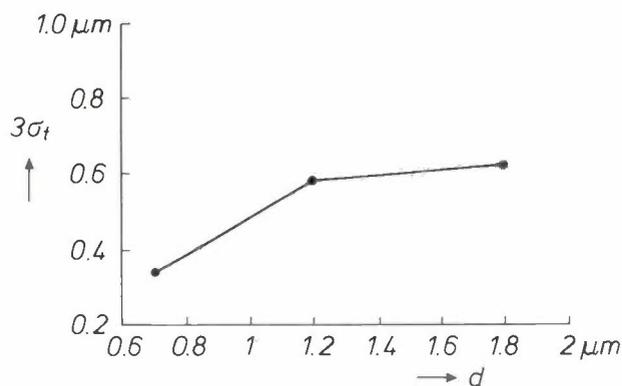


Fig. 8. The triple standard deviation $3\sigma_t$ as a function of detail size d . The standard deviation σ_t relates to a number of random faults that when taken together determine the error in the position of a line in relation to an adjacent contact plane after two masking steps.

Contacts between interconnection layers

In fig. 7 three of the four interconnection layers are visible, as well as a contact hole for a connection between layer *Al1* and a transistor electrode. The contact hole is formed by etching an opening into the SiO_2 , sputtering aluminium and etching away the surplus aluminium after a masking step.

Filling the holes in the oxide with aluminium becomes more difficult as the holes become smaller. For the 256-kbit SRAM the best method was to make the walls of the holes slope slightly. This approach will not be possible for later generations, such as the Mega memory, because the contact holes are smaller.

As the details in the integrated circuit become smaller, minute grooves in the SiO_2 give more and more trouble. Aluminium deposited in these grooves cannot always be removed by etching, and may introduce short-circuits between interconnection layers. One answer is to 'fill' the grooves in the oxide with an additional thin film of SiO_2 . This film can also reduce surface unevenness, thus increasing the accuracy of subsequent processing steps. The thin film of SiO_2 can be produced by spinning the wafer and spraying it with a colloidal orthosilicate solution, which is uniformly distributed over the surface by the spinning motion. The orthosilicates react so as to leave a thin film of SiO_2 behind, after heating. This film is called 'spin-on glass', and is designated *SOG* in fig. 7.

We shall now look at a number of problems that are of particular interest in scaling transistors down.

Shortening the transistor channel

In an NMOST (see fig. 4b) a potential difference of say +5 V between drain and source produces a depletion region at the drain. In this case a depletion region is one in which the surplus of holes in the p-type material has been removed by repulsion due to the positive voltage. The depletion occurs whether a voltage is applied to the gate or not. If the length of the n-type channel in the p-type material beneath the gate is reduced, the depletion region eventually touches the source. The potential barrier that prevents electrons from the source from penetrating into the p-type material then breaks down. The consequence is that the transistor conducts even when there is no voltage on the gate. This current leakage is known as 'punch-through'.

As the packing density of the integrated circuit increases, the shortening of the transistor channel

[5] S. Wittekoek, Optical aspects of the Silicon Repeater, Philips Tech. Rev. 41, 268-278, 1983/84; J. Biesterbos, A. Bouwer, G. van Engelen, G. van de Looij and J. van der Werf, A new lens for submicron lithography and its consequences for wafer stepper design, Proc. SPIE 633, 34-43, 1986.

means that this current leakage becomes increasingly significant. In SRAMs of very large capacity the resultant dissipation in all the transistors would cause overheating.

The answer is to position each NMOST in an island of increased boron doping. The higher hole concentration in such an island reduces the depletion region at the drain so that no contact is made with the source. For the free electrons in PMOSTs a similar effect is achieved by increased phosphorus doping in the n-type island for the transistor. Since the NMOSTs and the PMOSTs are then in separate islands, this modified CMOS technology is known as 'twin-tub CMOS'. The new 256-kbit SRAM is made in this technology, and it will also be used for the future Mega memory.

Another way of tackling the problems outlined above is to increase the threshold voltage, i.e. the gate voltage at which the transistor just starts to conduct. The threshold voltage of an NMOST can be increased by implanting additional B^+ ions immediately beneath the channel oxide, which is the oxide underneath the gate.

Hot electrons

Fig. 9a shows that the field in the y -direction (the lateral direction) can be as high as $300\,000\text{ V cm}^{-1}$ near the drain of a short-channel transistor. Electrons in the channel of an NMOST that have acquired high energy from the lateral field ('hot' electrons) can generate electron-hole pairs in collisions with the outer-shells of atoms. Some of the holes of the electron-hole pairs leak away to the back of the wafer where they cause a substrate current. Some of the electrons are deflected towards the gate. When they tunnel through the channel oxide, they generate a gate current. Since holes are also trapped in the channel oxide, the hot-electron effect causes a gradual change in the transistor characteristic, and therefore a short life.

The strength of the lateral field can be reduced by designing the NMOST as an LDD transistor (LDD stands for lightly doped drain), see fig. 9b^[6]. On the channel side of the source and drain an LDD transistor has regions with a phosphorus doping lower than the arsenic doping of the actual electrode region. The field-strength can therefore be almost halved. The SiO_2 spacers on either side of the gate are used to offset the n^+ regions. The lateral field-strength is high because we have kept to the unofficially standardized supply voltage of 5 V, in spite of the continual reduction in transistor dimensions. It is therefore not at all certain that it will be possible to keep to a supply voltage of 5 V for future integrated circuits in submicron technology.

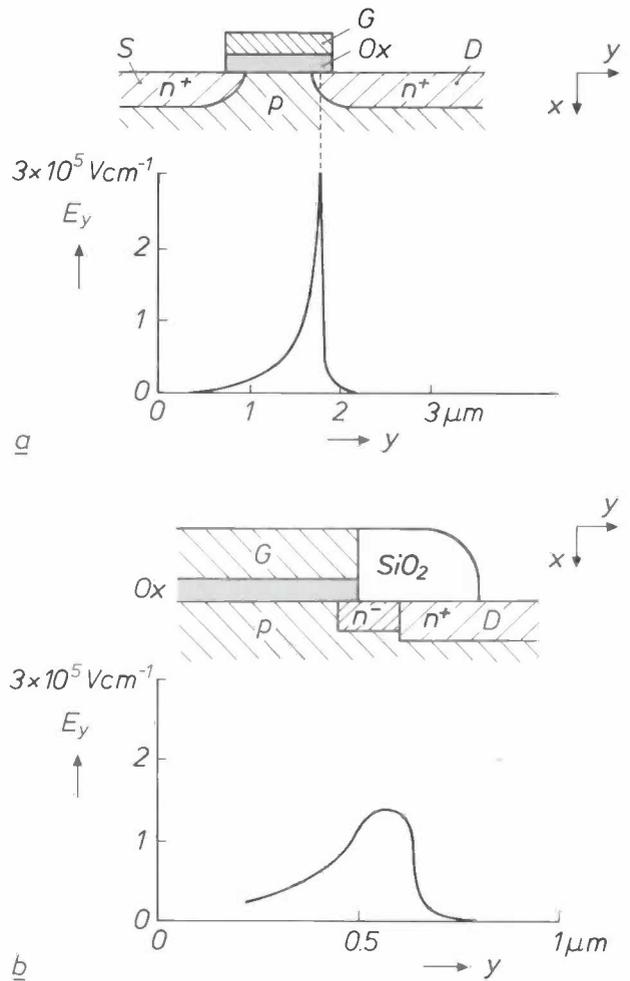


Fig. 9. a) Lateral field-strength E_y as a function of the coordinate y for a short-channel transistor. Electrons can acquire a high energy because of the high field-strength: these are 'hot' electrons. The abbreviations are explained in the caption to fig. 4. Ox channel oxide. b) The use of LDD transistors (LDD stands for lightly-doped drain) has the effect of almost halving the maximum of E_y . n^- regions on both sides of the channel oxide with a slightly lower concentration of free electrons than in the source and drain. (The SiO_2 spacers on both sides of the gate are due to the modified technology employed.) The curves in (a) and (b) are the results of computer simulations.

Narrowing the transistor channel

The 'bird's-beak' effect is the main cause of problems arising from narrowing of the transistor channel. We shall therefore first explain how this effect comes about.

At the start of the IC fabrication process the field oxide is formed (see fig. 4b) by the LOCOS technique

[6] P. T. J. Biermans and T. Poorter, Key parameters in the design of reliable submicron LDD devices, Proc. ESSDERC, Cambridge 1986, pp. 183-184.

[7] P. A. van der Plas, W. C. E. Snels, A. Stolmeijer, H. J. den Blanken and R. de Werdt, Field isolation process for submicron CMOS, Proc. Symp. on VLSI Technol., Karuizawa 1987, pp. 19-20.

mentioned above. In this technique regions with the required dimensions are etched into a thin film of Si_3N_4 on the wafer, and the wafer is then subjected to an oxidation process in which the nitride is used as an oxidation shield. Diffusions are made later through openings in the field oxide to form the sources and drains. In the oxidation process the SiO_2 grows into the wafer by about the same amount as on the wafer, so that the wafer remains comparatively flat. This is an advantage because steps in the surface could later cause line-width variations^[6].

To ensure good insulation of the individual transistors, the field-oxide layer must be relatively thick. The compound to form the oxygen for the oxidation should therefore be able to diffuse easily through the oxide. At the oxidation temperature the SiO_2 , which has vitreous properties, deforms fairly easily whereas the nitride does not. For this reason a thin oxide stress-relieving layer, *SRL*, is applied before the application of the nitride to reduce the surface stresses during the oxidation; see *fig. 10*. Without this layer crystal defects would form at the silicon surface. However, since the compound to form the oxygen for the oxidation diffuses easily through the oxide, a

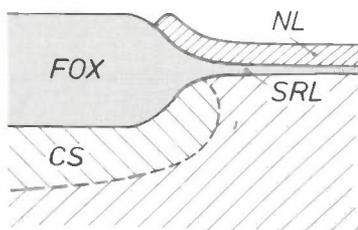


Fig. 10. The 'bird's-beak' effect. *FOX* field oxide. *NL* silicon-nitride layer, which acts as a mask during the oxidation step. *SRL* stress-relieving layer. *CS* channel stopper. The field oxide grows underneath the layer *NL* in the shape of a 'bird's beak'.

'wedge' of field oxide grows under the nitride, forcing it up slightly. This is known as the 'bird's-beak' effect.

In *fig. 10* another dopant layer, *CS*, is shown beneath the field oxide. This is a layer with extra doping to increase the threshold voltage of the parasitic transistor formed between adjacent transistors via the substrate. The extra-doped layer is called a 'channel stopper'.

The 'bird's beak' reduces the effective width of the transistor channel. With decreasing channel width the channel stoppers come very close together and may even overlap. Because of these effects the threshold voltage, which does not depend on the channel width for normal dimensions, becomes too high at extremely small channel widths. This gives rise to problems when the packing density of the circuit is increased. It is therefore necessary to suppress the bird's-beak effect as far as practicable.

The bird's beak can be suppressed by using silicon oxynitride, SiO_xN_y , instead of pure silicon dioxide for the stress-relieving layer^[7]. Silicon oxynitride reduces diffusion of the oxygen-supplying compound below the nitride, though it has the minor disadvantage that it deforms somewhat less easily.

It has been found that complete suppression of the bird's beak results in the formation of a groove at the edge of the field oxide when the nitride layer and the stress-relieving layer are etched away. The groove can give problems in subsequent anisotropic etching steps. However, a particular combination of *x* and *y* in SiO_xN_y can be found that just stops short of groove formation and only gives a small field-oxide wedge that is not a problem. *Fig. 11* shows a cross-section through the silicon surface at the start of the fabrication process, with a full bird's beak on the left and a suppressed one on the right.

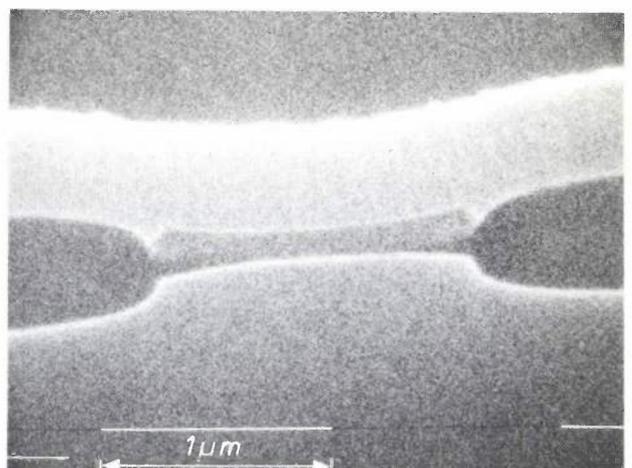
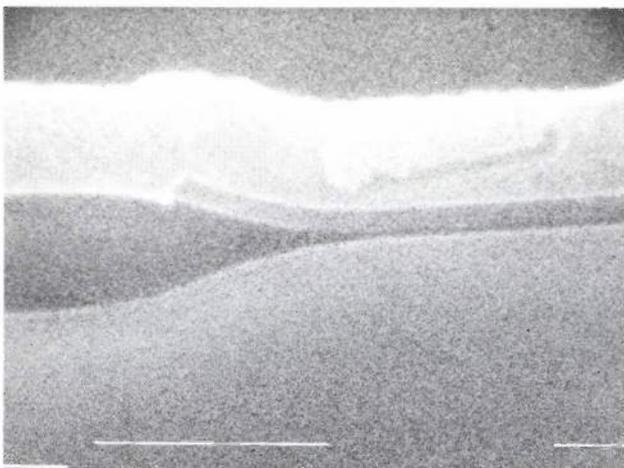


Fig. 11. Micrographs made with a Philips scanning electron microscope of a cross-section of a transistor channel, on the left with a bird's beak, on the right with a suppressed bird's beak.

Another way of reducing the threshold voltage, in addition to suppressing the bird's beak, is to use a lower temperature for the oxidation. This minimizes the diffusion of the channel stopper into the substrate.

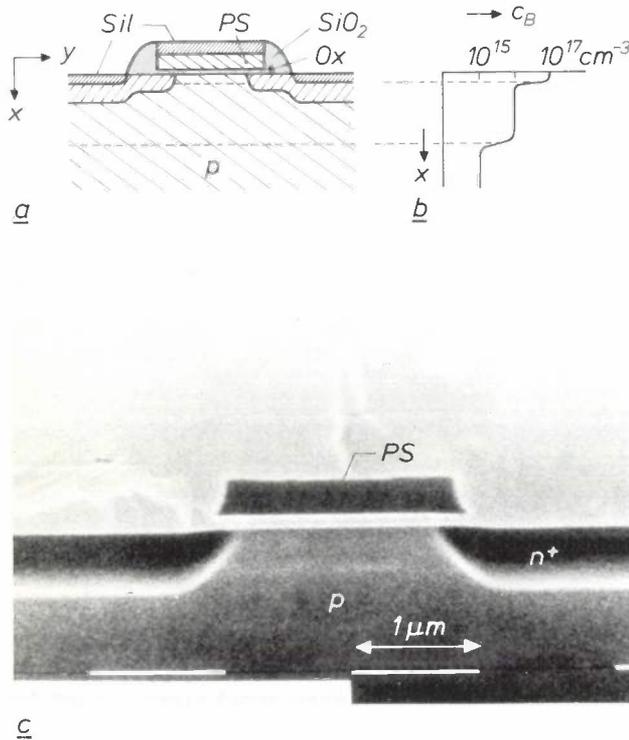


Fig. 12. *a*) Transistor (NMOST) incorporating various technological refinements. These refinements were necessary to suppress undesirable effects due to scaling-down of the details. *Sil* silicide. The other abbreviations have been explained in the captions to previous figures. *b*) Concentration c_B of boron atoms as a function of the coordinate x . *c*) Conventional transistor with 'normal' dimensions, without all these refinements.

Resistance of the electrodes

In a scaled-down transistor the diffused regions for the source and drain are shallower. This means that their resistance is higher. The same applies to the gate, and the result is that the speed of the circuit is limited. The electrodes in the Mega memory will therefore be shunted by a layer of silicide, which has a very low resistivity.

It will be clear from what we have said above that the production of a transistor with a channel length and width of about $1 \mu\text{m}$ requires extremely accurate control of the IC-fabrication processes. *Fig. 12a* shows a diagram of an NMOST incorporating all the refinements mentioned in this article. *Fig. 12b* shows the concentration of boron atoms in the p-type material beneath the gate. The LDD transistor in *fig. 12a* is less than half the size of the conventional transistor of *fig. 12c* without all these extras.

Summary. The static random-access memory (SRAM) with a capacity of 256 kbit, developed at Philips Research Laboratories, contains details of $1.2 \mu\text{m}$. The cells in this memory, made in full CMOS technology (complementary metal-oxide semiconductor), consist of four transistors, plus two connecting transistors. The current requirement and access time of the SRAM are both small, because of the application of the 'dynamic double-word-line technique'. Scaling down the details in an integrated circuit entails various problems. Shortening the transistor channel causes the current leakage known as 'punch-through' and premature failure of the transistor due to 'hot' electrons. Narrowing the transistor channel leads to difficulties because of the 'bird's-beak' effect, which raises the threshold voltage. These problems have been solved by adapting the technology, including the change to LDD transistors (LDD: lightly doped drain). Other problems can arise from contacts between the interconnection layers and the resistance of the conductors.

1938

THEN AND NOW

1988

Plastics in vacuum cleaners

Phenolformaldehyde resin used to be known as 'Philite' at Philips, since Bakelite was a trademark at the time. Invented in 1905 by Leo Hendrik Baekeland, an American of Belgian origin, it was one of the first plastics, and Philips used it extensively for parts like vacuum-cleaner components because of its 'mechanical, aesthetic and electrical properties'; see the photograph at the upper left^[*]. Philite is a 'thermosetting resin': the material sets permanently in the mould as the result of a chemical reaction and is only plastic when the mould is filled.

Today almost all the plastic components for vacuum cleaners are made of thermoplastics: these soften on heating, and are therefore easy to process. Nowadays there is a wide range of thermoplastics, so that the designer has a much freer hand than his counterpart in 1938. Only two colours of Philite were available, for example — black or brown. If the designer wanted another colour, the component had to be painted.

The photograph at the upper right shows the interior of a Philips vacuum cleaner that recently went into



production. The plastic components have been partly cut away. The dust bag fits into the large space on the right. The air drawn in is filtered by the dust bag, and the airstream passes through a second filter, the fan, the motor and a silencer. This is a 'reactive silencer'^[**]: and is formed by the hollow spaces in the top and bottom of the cleaner. The thick ring of mineral wool around the motor and the sealed casing of polypropylene — a thermoplastic — enclosing the motor-fan assembly help to make the machine extremely quiet.

The photograph below shows the complete vacuum cleaner. The combination of a smoothly rounded outer wall, uniformly coloured throughout, with complicated inner spaces can only be obtained with modern thermoplastics. In 1938 no-one dared to

dream of such an elegant and quiet vacuum cleaner with high suction power, swivel-top house, auto cord rewind, sturdy running wheels, spring-loaded swivel wheels, electronically controlled air flow and a clear indication of when the large paper dust bag is full.



[*] From: Philips Technical Review, January 1938.

[**] J. Crucq, Theory and practice of noise control in equipment, Philips Tech. Rev. 44, No. 4, Summer 1988.

EXPERTISE: an expert system for infrared spectrum evaluation

T. Blaffert

There is no doubt that the most important aid for the modern researcher is the computer. It will perform many tasks that people would find monotonous and time-consuming, and it will do them better and faster. But the tasks that the computer can perform are generally the routine ones. Thinking and reasoning still have to be done by people. Even 'expert systems', which are intended to stand in for specialists in a particular field, are really only 'tools' that are quick, convenient, and absolutely consistent.

Introduction

As once suggested in a previous issue of this journal^[1], analytical chemistry is an important branch of applied science, but perhaps not the strongest in imaginative appeal. It is often necessary, especially in organic chemistry, to characterize a wide variety of compounds that often differ little from one another, and analysis by conventional methods is a tedious and lengthy process. However, now that physical methods have become established in analytical chemistry and data is produced and processed in digital form on the computer, the problem can be tackled in a quicker and more imaginative way. Computer programs known as 'expert systems' can simplify and speed up the work of the specialist, who will always have his part to play.

In essence expert systems are programs that can access a 'knowledge base' built up by specialists in a particular field. They are being developed in various branches of science, in industry and in the financial world, but most of them are still only at the experi-

mental stage. An expert system for chemical analysis can identify unknown materials by comparing data with similar data of known origin. To some extent it can 'reason' like the specialist. The use of expert systems makes the knowledge amassed by the specialists more readily available at any time or place.

One of the first expert systems for chemical analysis was DENDRAL. This system has been under development at Stanford University since the sixties^[2]. DENDRAL makes use of results from mass spectrography to track down the possible molecular structures of compounds of known atoms. Other expert systems followed that had similar objectives to DENDRAL, but used spectrographic data of a different kind, such as Magnetic Resonance or infrared spectra. The infrared spectra are very useful for identifying organic molecules, because they contain a large amount of information and can be used as 'fingerprints'. This article will describe EXPERTISE, an EXPERT system for Infrared Spectrum Evaluation. EXPERTISE identifies chemical compounds and structures from their infrared spectra.

Dipl.-Phys. T. Blaffert is with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

An infrared spectrum is obtained when a sample absorbs radiation in the infrared part of the electromagnetic spectrum (with wave numbers between 4000 and 650 cm^{-1}). The absorbed radiation causes a transition between two energy levels in a molecule. These energy levels are associated with the vibrations of the atoms in the molecule. It is clear that the vibrations of the atoms in a molecule will be greatly affected by the nature and position of the surrounding atoms, i.e. by the structure of the molecule. This is why infrared spectra can be used as 'fingerprints' for characterizing organic substances or to throw some light on their molecular structure. Fig. 1 shows the spectra of toluene ($\text{C}_6\text{H}_5\text{CH}_3$) and aniline ($\text{C}_6\text{H}_5\text{NH}_2$). Similarities in the structures of these compounds (the phenyl groups: C_6H_5-) produce peaks at the same positions, whereas differences (the NH_2 and the CH_3 groups) produce peaks at different positions.

One way of using an expert system to analyse an unknown substance would be to compare its infrared spectrum with as many spectra as possible. But since there are more than seven million known organic compounds, it is rather difficult to obtain an infrared spectrum for each of them, quite apart from the problems of storing all this information in a knowledge base. Another problem is that newly synthesized substances will not be recognized. On the other hand an experienced analytical chemist would realize that parts of the spectrum were due to the presence of particular substructures (functional groups) in the molecule. By combining the pieces of information about the substructures he can eventually decide on the structure of the molecule. If this aspect of the working methods of the specialist can be incorporated into an expert system, it will greatly add to the practical value of the system.

An expert system that will work in this way must have several distinct phases in its program. The program must be able to:

- subdivide known structural formulae into substructures, i.e. it must be able to break down the structural formulae in the knowledge base into appropriate smaller units,
- recognize spectral features characteristic of the substructures it has generated,
- search for substructures by comparing parts of unknown spectra with spectral features present in the database,
- generate structural formulae from the substructures it has recognized.

These were the requirements that led us to develop EXPERTISE.

The operation of the system can be briefly explained as follows. First of all, a number of compounds are chosen as the nucleus for the knowledge base, and their

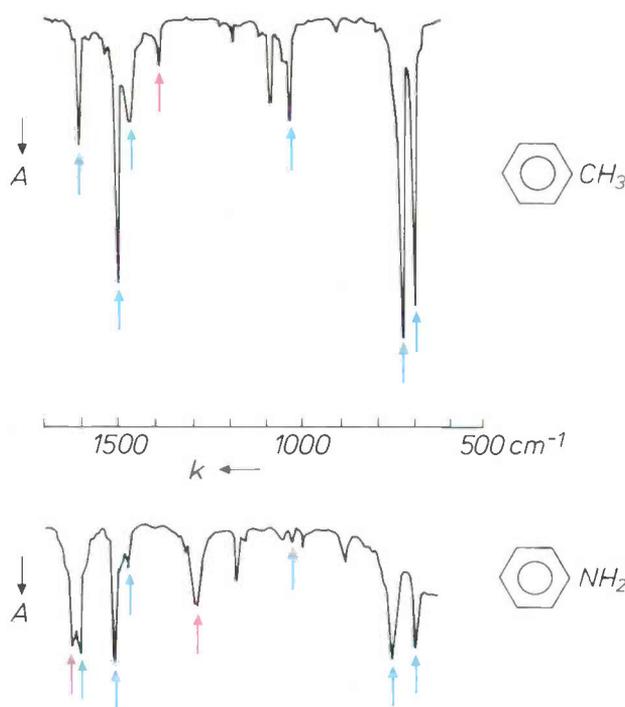


Fig. 1. Infrared spectra (absorbance A plotted against wave number k) for toluene and aniline, compounds that both contain a phenyl group. Similarities in the spectra — absorbance peaks at the same wave number — are indicated by blue arrows. These peaks are due to the absorbance by the phenyl group. Other peaks in the spectra, indicated by red arrows, are due to absorbance by the CH_3 group (in toluene) and the NH_2 group (in aniline).

structural formulae, infrared spectra and formulae and spectral features of substructures are all stored in the knowledge base. When the content of the knowledge base has reached a sizeable value, an infrared spectrum of unknown origin can be applied to the system, as input, to generate a structural formula from the recognition of substructures. The system can also search for the complete structural formula without the user having to divide up the spectrum into subspectra. This function of the system is a library search procedure and can be used in the analysis of mixtures. We shall not consider this aspect of the system further here^[3]. The four other characteristics of the system — substructure generation, feature learning, substructure search and structure generation — will be discussed in the following sections. Before going into details of the operation of the system let us first look at a general outline of the entire system and the relationships between its various component parts.

[1] W. F. Knippenberg, *Inorganic chemical analysis*, Philips Tech. Rev. 34, 298-304, 1974.

[2] B. G. Buchanan and E. A. Feigenbaum, *Dendral and meta-dendral*, *Artif. Intell.* 11, 5-24, 1978.

[3] T. Blaffert, *Computer-assisted multicomponent spectral analysis with fuzzy data sets*, *Anal. Chim. Acta* 161, 135-148, 1984.

The design of the system

The central part of an expert system is the knowledge base to which all operations have permanent access. In all the systems the answer to the problem stated is obtained by comparing unidentified data with data of known origin. The comparison is made with varying degrees of ingenuity in the different systems, but in principle this comparison is the fundamental operation in all of them.

Fig. 2 shows a schematic representation of EXPERTISE. There are two possible inputs to the system: an infrared spectrum of a sample of known chemical composition and structure and an infrared spectrum of an unidentified sample. Let us first consider what the system does with the first kind of input. The infrared spectrum and the chemical structural formula are stored in the spectrum library (I). The structural formula is divided into substructures (2) and the charac-

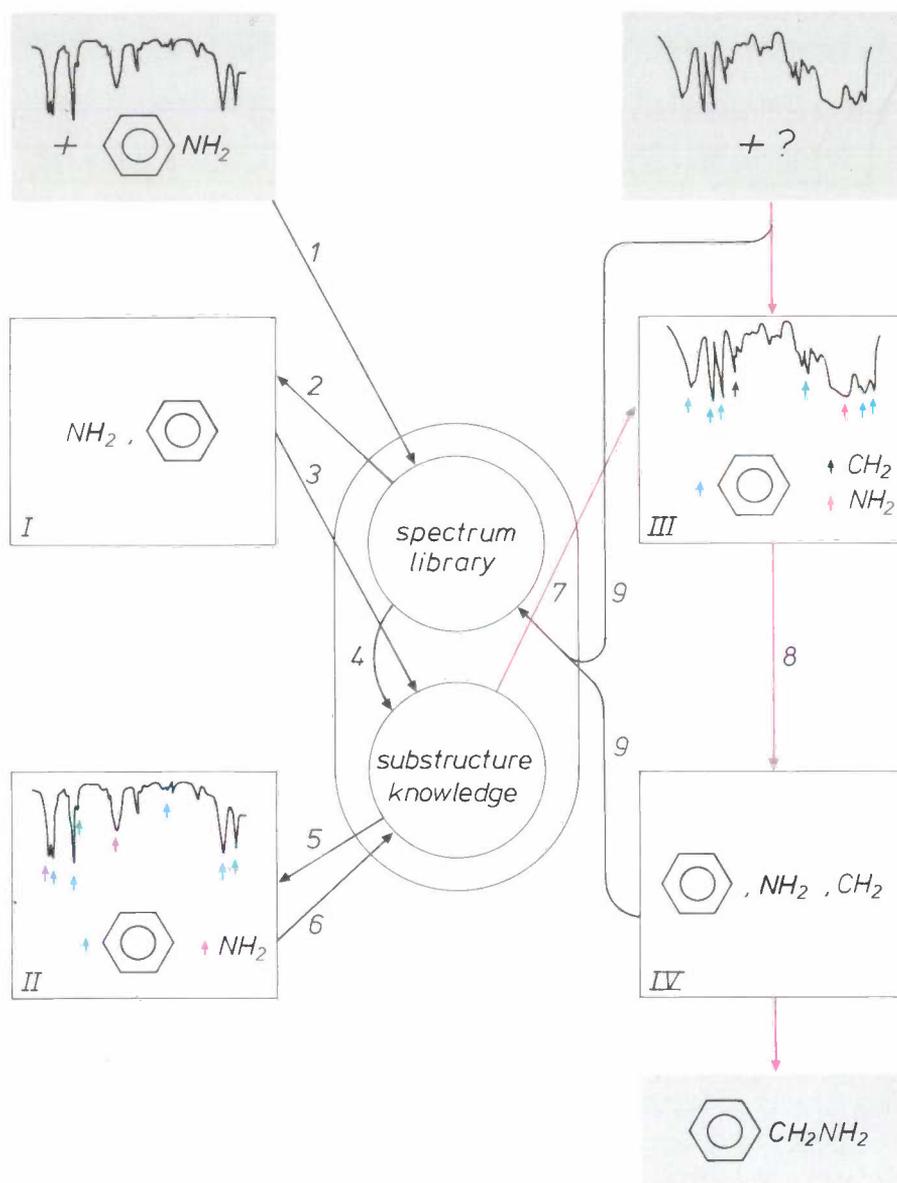


Fig. 2. A general outline of the EXPERTISE system. The central part of the figure represents the knowledge base, which contains the information about chemical compounds (spectra and structural formulae) for complete molecules and parts of molecules (functional groups). The left-hand half of the figure gives an impression of the operations performed in assembling the data for the knowledge base, while the right-hand half illustrates the actual application of the system for the identification of unknown compounds. The inputs to the system are infrared spectra. In assembling the data for the knowledge base the structural formulae of the compounds must also be provided as input. The output of the system consists of one or more possible structural formulae for the unknown compound.

teristic features in the spectra of structures containing these substructures are 'learned' (4,5). Both substructure and corresponding features are stored in the substructure knowledge base (3,6). The spectrum library and the substructure knowledge base form the database of the system.

The second kind of input, a spectrum of an unknown sample, is treated in a different way. First of all, spectral features from the knowledge base are used to determine which substructures are present in the sample (7). The structural formula of the unknown sample is then generated from these substructures (8). The result is stored in the knowledge base, along with the input, the spectrum (9).

In the schematic representation of fig. 2 two different routes can be seen. The first relates to the operations that are performed when the input consists of the spectrum and the corresponding structural formula, and is used for the acquisition of the knowledge. The second route is taken when the knowledge base is used for identifying unknown compounds. These two main functions of the system will be dealt with in the following sections.

Knowledge acquisition

To identify the spectral features associated with the substructures, it is first of all necessary to subdivide the known structures into substructures. This must be done in such a way that there is only a limited number of ways of combining the substructures to form a structure. The nature of the bond between the different atoms must therefore be considered as well.

The way in which a chemical structure is divided into substructures is illustrated in fig. 3. The structural formula of $\text{CH}_3\text{CH}_2\text{NH}_2$, as shown in fig. 3a, contains two kinds of symbols: circles for the atoms from which the molecule is built up, and rectangles for the bonds between the atoms. This representation is used to define 'superelements' and 'superbonds': superbonds are defined as bonds that do not have a terminal atom plus the attached elements and their bonds (indicated by lower-case letters, *a, b*), superelements are the parts of the structure that are linked by superbonds (indicated by capitals, *A, B*). In terms of superbonds and superelements the structural formula of $\text{CH}_3\text{CH}_2\text{NH}_2$ is $A-a-B-b-C$ (fig. 3c). The superelements together with the superbonds are called 'superatoms'. They are substructures of the type used in EXPERTISE. The effect of neighbouring atoms on the substructure is taken into account by including the superbonds in the substructure. This part of the program is indicated by *I* in fig. 2.

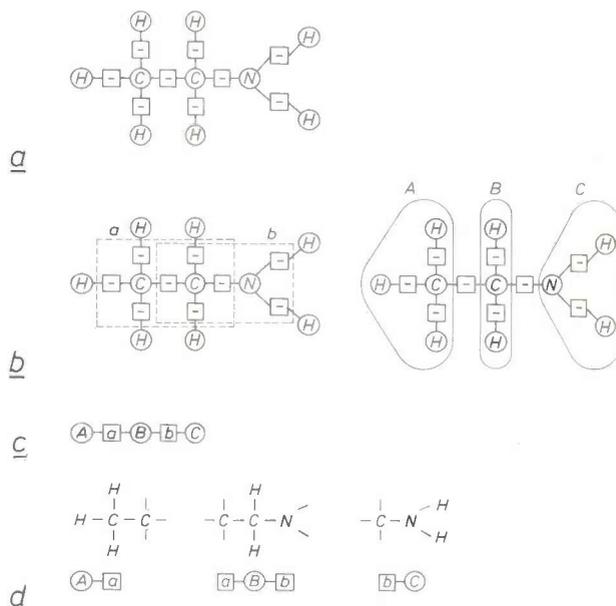


Fig. 3. Subdivision of a structure into substructures. *a*) The structure is represented schematically with the aid of two symbols. Chemical elements are shown as circles, bonds between elements as rectangles. *b*) Superbonds and superelements are derived from this representation. Bonds with no terminal atom connected to them form the centres of superbonds. These are defined as bonds plus the attached atoms and their bonds (indicated by dashed lines and lower-case letters). The central bonds separate the structure into superelements (indicated by continuous lines and capital letters). *c*) A representation of the chemical structure in terms of superbonds and superelements is called a superstructure. *d*) The substructures used in EXPERTISE are the superelements plus all the attached superbonds, called superatoms. The features of the substructures are 'learned' and will be recognized in an unknown spectrum.

The second step necessary for the construction of the knowledge base is finding the spectral features associated with the substructures. A large number of compounds that contain a particular substructure are therefore selected. For example, to search for the spectral features of the substructure *C* (NH_2) of fig. 3, the infrared spectra of a large number of compounds that contain this group are used to search for similar characteristics. These characteristics are then separated from the rest of the spectra and stored in the knowledge base. We shall now look at the details of this procedure.

The parameters by which we recognize a spectral feature are the wave number and the intensity of an absorption peak. These parameters are found to be the most important characteristics for distinguishing between the various spectral features. Spectral features are therefore represented by points in the (wave number)-(intensity) plane. Since the wave number and the intensity both depend to some extent on the environment of the substructure, the representation consists of 'intervals' rather than single points. To deter-

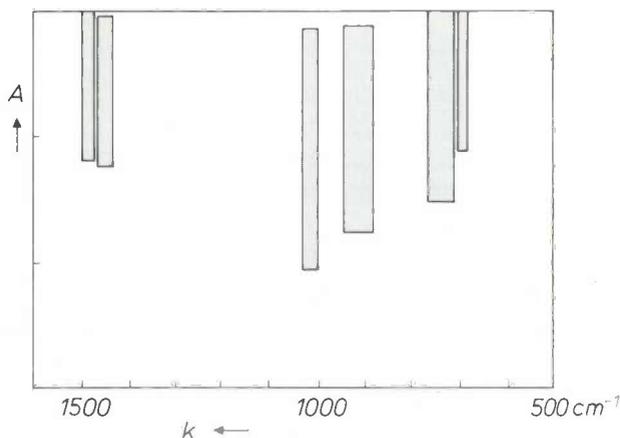


Fig. 4. Learning the characteristic features of a substructure. This graph shows the absorbance plotted against the wave number for all the peaks in the infrared spectra of a number of compounds that all contain a phenyl group, and is used to determine the spectral features of this phenyl group. The intervals that contain the peaks of all the compounds are found, and the intervals with the smallest areas are chosen. If an unknown infrared spectrum has peaks in each of these intervals, then the phenyl group is clearly part of the structure of the unknown substance.

mine the spectral characteristics of the phenyl group we use a graph of the peak intensity as a function of the wave number for a number of compounds that contain phenyl groups. The intervals with the smallest possible areas where peaks for all these substances occur now form the spectral characteristics. In *fig. 4* these are indicated by the shaded areas. The peaks not included in these areas must be associated with other substructures, and not with the phenyl groups. The intervals, together with the substructure we are looking for, the phenyl group, are stored in the knowledge base and can be used later for identification. It can be seen from *fig. 4* that the intervals are narrow and high. This what we would expect, of course, since the intensity is more dependent than the wave number on properties of the sample that are unrelated to the structure, such as the quantity. This aspect of the system is indicated in *fig. 2* by *II*.

Application of the stored knowledge

In this section we shall consider the procedure followed when the input to the system is an infrared spectrum of unknown origin (*fig. 2*). First of all the possible substructures must be identified from these spectra and then they must be combined to obtain the actual structure.

To identify the substructure we compare the wave number and the intensity of the absorption peaks in the spectrum of unknown origin with the spectral features

of the substructures, i.e. the intervals in the (wave number)-(intensity) plane, in the knowledge base. If all the intervals that form a certain substructural feature contain at least one peak, then we know that this substructure is part of the unknown structure. This is a very simple and straightforward rule. There is one problem, however. If, say as a result of temperature fluctuations, the peak parameters of the unknown spectrum fall just outside the intervals that characterize a substructure, this substructure may not be recognized. This deficiency in the system can be corrected by assigning a 'membership function' f to the intervals. This function, which can have a value between 0 and 1, is used to take account of the distance between the position of the peak and the interval. Its value indicates the extent to which the peak can be considered as part of the interval; in other words it indicates the degree of recognition of a substructure. If the peak is inside the interval, the function has a value of 1. If the peak is clearly outside the interval, the value is zero, and in the intermediate positions its value decreases continuously from 1 to 0. The boundaries of the interval are 'fuzzy'. This method is often used in computer science, and is widely known as 'fuzzy-set theory'.

In *fig. 5* the value of the function f is shown for the intervals of *fig. 4*, the spectral features of the phenyl group. Peak P_1 is inside the interval, so the value of the function is 1. Peak P_2 is on the fuzzy boundary, with a value of say 0.5. Every peak in the spectrum of the unknown compound is examined to see if it comes within an interval or belongs in some degree to the

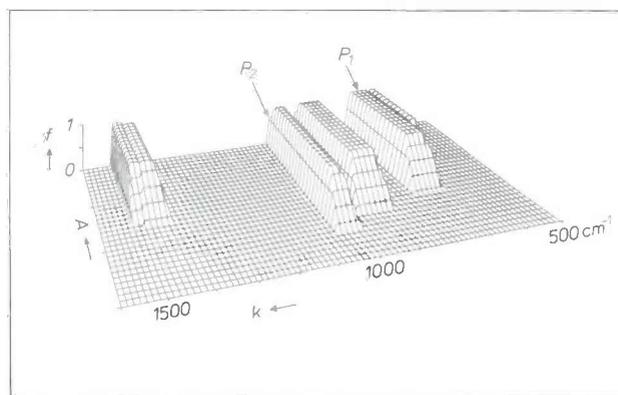


Fig. 5. Schematic representation of the intervals of *fig. 4* with 'fuzzy' boundaries. The position of the intervals is plotted in the (intensity)-(wave number) plane, and the value of the function f is plotted along the third axis. Its value is 1 inside the interval and gradually decreases to zero outside the interval. Representations of this type are used in searching for substructures. The boundaries of the intervals are 'fuzzified' to prevent substructures from being rejected because of slight changes in the location of the peaks, e.g. because of statistical variations. If peaks of an infrared spectrum are found in all these intervals or on the fuzzy boundaries of the intervals, the phenyl group is part of the structure.

nearest interval. The system then collects the substructures for which all the peaks of the input spectrum can be fully assigned to the corresponding intervals, with the value 1 for the membership function. If this set of substructures does not give a complete structure (see below) or if the user would like alternative solutions, the system can define another set of substructures. In this new set more of the peaks can be assigned to intervals, because a membership value below 1 but above a certain threshold value M_t is allowed. The previous set of substructures is included in this new set, of course. Substructures found in this way can be thought of as belonging to the structure to a degree indicated by the threshold value of the function f . This is an extra parameter that has to be taken into account in deriving the structure from the substructures discovered. The process of reducing the threshold value and defining the corresponding set of substructures goes on till a complete structure can be generated from this set or the user does not want any more alternative solutions. This aspect of the system, the recognition of substructures, is indicated by III in fig. 2.

To complete the description of the system it is still necessary to explain how a structural formula for the unknown material is generated from the set of substructures discovered. It is not immediately clear whether all the substructures (with different threshold values M_t) should be used in the process. The system therefore generates a number of possible arrangements for the total structure, if necessary. It also gives an extra output that can be used for ranking in the form of the minimum value M_t for all the substructures used. The output of the system can therefore contain more than one structure and it is left to the chemist using the system to decide which is the most likely structure. In reaching this decision the chemist may find it helpful to take account of the threshold values, and of course he may have access to extra information not in the system. Although the system does not give a unique answer, it does have two great advantages. First, the system does not overlook any structures, and secondly, all the possible structures are found fairly quickly (more quickly than the chemist could find them). We shall now look more closely at the process for generating the structure from these sets of substructures (see IV in fig. 2).

Let us assume that we have the set of substructures (superatoms in the terminology of the system); $A-a-$, $-a-B-b-$, $-b-C$, $-d-C$ (see fig. 6). The system sets to work in the following way. Superatoms can be linked together by saturating their open superbonds with the identical open superbonds of other superatoms. This linking can continue until there are

no more open superbonds in the superstructure thus generated. The chemical structure is then obtained by simply inserting the chemical substructural formula for the superelement and the bond type for the superbond into the superstructure.

In the example of fig. 6 this is done in the following way. The last superatom $-d-C$ is deleted, since there is no other superelement with the superbond $-d-$ for it to combine with. One of the remaining superatoms is now chosen to start with, e.g. $A-a-$. This can be combined with another superelement with the superbond $-a$. Two choices are possible here. The combination $A-a-A$ is not accepted because not all of the superatoms in the list have been used. The other possibility, $A-a-B-b-$, is not a complete structure and can in turn be combined with $-b-B-a-$ or $-b-C$. The first possibility is not accepted because a non-terminal superatom has been used twice. This constraint is imposed to prevent the generation of endless chains. In evaluating the results, the chemist has to remember that in reality some substructures

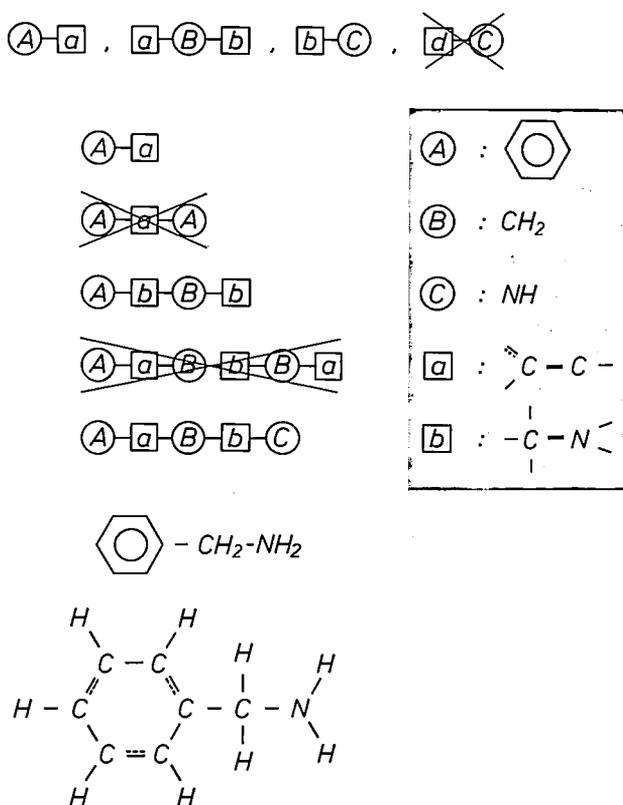


Fig. 6. Generation of the structure from the substructures found in the substructure search (listed at the top). The last substructure on the list is deleted since there is no other substructure with the same bond. The substructures are combined and eventually the structure $A-a-B-b-C$ is found. (See text for the rejection of other possibilities.) The symbols that represent the superatoms and superbonds are replaced by the structural formulae that they represent. The output of the system consists of these chemical structures with the corresponding values of M_t .

may occur more than once in the structure. The second possibility results in a complete superstructure and is considered to be the correct solution. The knowledge base makes the substitution for the superbonds and the superelements, giving $C_6H_5 - CH_2 - NH_2$. In this case the output of the system consists of a complete structure. In more complex situations, where decisions like the ones above are not possible, the system may offer more possible structures together with the parameter M_t mentioned earlier.

At present the system is still in an experimental phase, but we are already able to quote a few initial results that give an indication of its performance. About 95% of the substructures are identified correctly, but the system may offer some 10 to 20% more substructures than are actually present. It is clear that the analytical chemist still has an essential contribution to make here. Another feature of the system is that it can distinguish between disubstituted benzene

derivatives that differ only in the relative positions of the functional groups (*ortho*-, *meta*- and *para*- positions). A complete evaluation of the system is not as yet available, however.

Summary. Infrared spectra are exceptionally useful for providing information about chemical structure. Analytical chemists who use these spectra to identify the chemical structure of unknown or newly synthesized compounds will find the EXPERTISE expert system a great help. Besides identifying complete spectra by comparing them with the spectra in an extensive library, this system can to some extent 'reason' as the chemist does. This is achieved by subdividing the structures in the library into substructures, establishing the spectral features related to these substructures, recognizing these features in spectra of unidentified compounds and combining the substructures to form a complete structure. The procedure for identifying the substructures from the spectral features includes a safety factor to prevent substructures from being excluded in error because of statistical variations in the spectra. This is done with the aid of fuzzy-set theory. The system may provide more than one result, in which case the chemist has to make the final decision. The great advantage of such systems is that all the possible results are found relatively quickly.

Future trends in optical recording

G. E. Thomas

The text of the article below is largely taken from the invited paper that the author presented at the International Symposium on Optical Memory 1987, held in Tokyo from the 16th to the 18th of September 1987. In this article he describes the most important trends that can be recognized today in the field of optical recording. Besides the well-known audio Compact Disc ('CD'), the LaserVision disc and the Digital Optical Recording ('DOR') disc for data storage, there will be new applications in the future, especially if erasable optical recording becomes a practical reality.

Introduction

It is undoubtedly only a coincidence that the symposium for which this material was originally put together coincided with the fifteenth anniversary of the first public demonstration of optical recording, in September 1972 at Philips Research Laboratories in Eindhoven. To resurrect a cliché, it will be clear to all who have been involved in optical recording^[1] that the field has been characterized during its brief history by turbulent and, at times, explosive growth. Industrially, the field of optical recording is a curious mixture of extremely successful consumer applications, such as the video disc and — in particular — the Compact Disc, and professional data-storage applications in which the major breakthrough in the market is avidly predicted but still awaited. The word 'curious' has been deliberately chosen here, since the two outstanding advantages of optical disc recording — the high storage capacity per unit area and the random-access capabilities — which could be expected to ensure its breakthrough in the professional area, have not yet led to that goal. On the other hand, these advantages have been quite successfully exploited in the products for consumer entertainment. Against this background, it is clear that it is very difficult to pre-

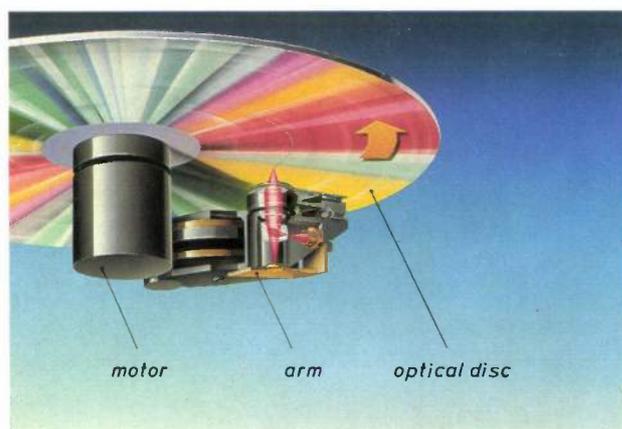


Fig. 1. An optical recording system consists of three main components: the optical disc, the arm and the motor. The information on the optical disc is contained in microscopically small optical details arranged on the disc in a spiral path, rather as in a conventional audio disc. These details are 'scanned' by a beam of light, from a semiconductor laser, which is reflected from the surface of the disc on to a detector. The arm contains all the facilities necessary for the light beam to follow the information track in the radial direction and for focusing the beam on to exactly the right position. The motor must always rotate the disc at the correct speed; in most optical recording systems this speed depends on the radial position of the arm. The unit containing the optical components is usually referred to as the 'light pen' of the system.

[1] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen and K. Schouhamer Immink, *Principles of Optical Disc Systems*, Adam Hilger, Bristol 1985.

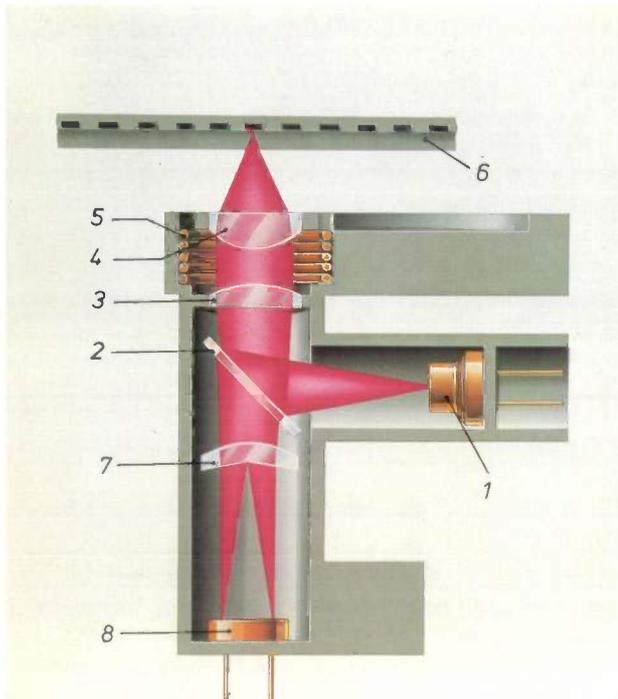


Fig. 2. Schematic drawing showing the most important components of the 'light pen' in an optical recording system. 1 semiconductor laser. 2 half-silvered mirror. 3 collimator lens, which makes the beam parallel. 4 objective lens for focusing the beam. 5 actuator, which determines the position of the objective lens. 6 optical disc. 7 beam-splitter. 8 detector for the beam reflected from the disc. The detector is in fact a multiple sensing device, and information about focusing and positioning of the beam with respect to the optical track on the disc is also obtained during the read-out. The variations in the light beam reflected from the disc can depend on various physical principles: differences in level ('pits') in a reflecting layer can give differences in intensity on reflection, differences in direction of magnetization of a recording film can affect the polarization of the reflected light ('magneto-optic recording'), or local differences in crystal structure can give differences in intensity on reflection ('phase-change recording').

dict future trends in optical recording. It is appropriate at this symposium to let the perceived requirements of sophisticated professional systems guide the predictions. However, it may well turn out that future adaptations of optical recording in the consumer area will provide much of the stimulus for new developments. In particular, the demands placed on storage systems for high-definition video signals are fairly stringent. There will certainly be a need for increased storage capacities and higher data-transfer rates.

As we look back on the brief history of optical recording systems, we can distinguish a number of research and development phases. The first phase was obviously concerned with investigating and defining the basic strategies to be followed in designing the optics, servomechanisms, coding and modulation methods, electronics and media for both analog (video)

and digital (CD) read-only systems^{[1]-[3]}. The following major surge in activity dealt with the development of media that were suitable for recording^{[4]-[6]} and the associated investigation of higher-power laser technology and modulation methods adapted to these media. The current widespread efforts in the technology of erasable recording (based on both magneto-optics and amorphous/crystalline phase-change strategies) represent the third major development phase in optical recording^{[7][8]}.

It is safe to predict that future developments will mainly centre on improvement of the existing systems, both in terms of a significantly higher storage density and of higher data rates. An improvement in the total performance assumes an improvement in the individual optical recording subsystems, from the laser to the disc itself; see *figs 1 and 2*.

Coding

Fig. 3 summarizes the current state of affairs with respect to storage density on optical media. In this figure, the bit density (in Mbit/cm²) for a number of current systems is plotted as a function of the spatial optical cut-off frequency ($f_c = 2NA/\lambda$) for the optics and wavelengths used. As can be seen from the cluster of points in the lower left-hand corner, only one area of this diagram has been exploited to any degree. It is clear that the current situation is determined by the availability of semiconductor lasers in the wavelength range from 720 to 820 nm and by the relative ease of mass-producing diffraction-limited optics with a numerical aperture of 0.4 to 0.5. The spread in bit density is a measure of the effectiveness of various modulation methods in exploiting the characteristics of the optical recording channel. It is also interesting to note that the current CD system^[2], in which pulse-length modulation is used, has already attained a respectable data density. If, in a hypothetical experiment, we go from the analog frequency-modulation methods like those used in video-disc systems (e.g. for PAL signals) to the digital domain via a conversion with 8 bits per pixel, we obtain a similar relatively high storage density. Both systems exploit the high accuracy with which the leading or trailing edges (or both) of optical effects on a disc can be located to encode the data stream.

Detailed investigations of the signal characteristics of the optical recording channel have only been made comparatively recently^[9]. Such investigations require sophisticated optical recorders and signal-processing facilities. The channel capacity is ultimately limited by noise. For the present, however, the capacity is mainly limited by intersymbol interference and crosstalk. As better quantitative measurement techniques become

available, so that the properties of various media (replicated, write-once and erasable) can be characterized in more detail, it is becoming increasingly probable that advanced coding and modulation strategies will lead to an increase in the bit density. This could amount to a factor of 4 to 5 with respect to the present situation.

An indication of some future trends is given in fig. 3. Starting with the characteristics of the CD system, for example, and assuming that noise is not the chief capacity-limiting factor of the channel, then the achievable bit density should increase with the square of the cut-off frequency. A number of reference lines (dashed) are shown for optics with a larger numerical aperture and for shorter recording wavelengths (possibly obtained by frequency doubling). The feasibility of the various combinations will be discussed briefly below.

Finally, fig. 3 shows the bit density resulting from a model calculation of the performance of the optical recording channel, based on the assumption that Gaussian white noise forms the ultimate limitation. This ultimate performance (it is proportional to $f_c^{3/2}$) is the theoretical limit for diffraction-limited optics. As can be seen, a considerable gap exists with respect to the current systems. It is unrealistic to presume that this gap can ever be completely eliminated, since practical considerations such as manufacturing tolerances and the like remain. Nevertheless, a combination of tactics (i.e. improving the coding techniques in combination with a higher f_c) will undoubtedly offer prospects of achieving recording densities that are higher by at least an order of magnitude.

Lasers

The future trends in the development of semiconductor lasers for optical recording are fairly easy to predict. A subject of great importance at the moment is the development of lasers in the wavelength range around 800 nm with a high useful power output. The term 'useful' here relates to the power in a beam of sufficiently high quality for application in the light pen of an optical recording system. The lifetime and the feedback characteristics of these lasers^[10] are also critical parameters. The trend is clearly towards the use of higher powers for advanced write-once and erasable materials. This is a natural consequence of the general desire to attain higher data rates in optical recording. Here, of course, developments in the field of laser arrays (see fig. 4) in which each laser can be independently modulated are eagerly awaited, since the possibilities of achieving higher data rates through the use of parallel recording are obvious.

A second hoped-for trend in lasers is the reduction of the wavelength. This can be achieved by making

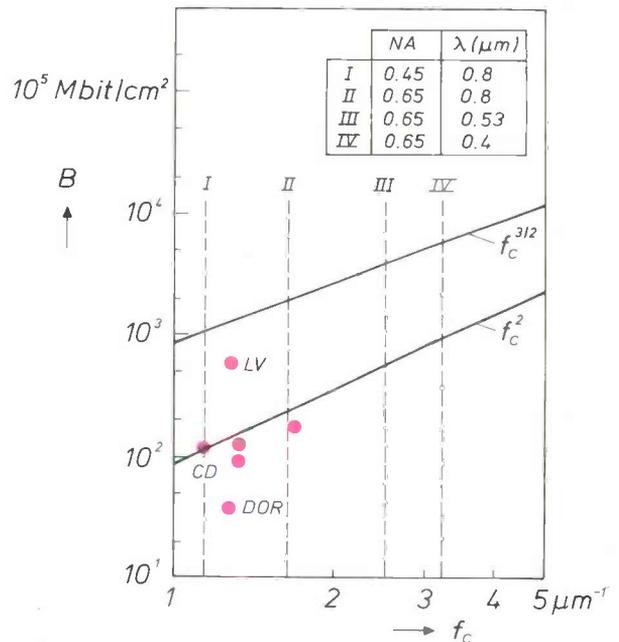


Fig. 3. Storage density B as a function of the spatial optical cut-off frequency $f_c = 2NA/\lambda$, where NA is the numerical aperture and λ is the wavelength of the light (in μm). The points represent the current state of the technology for a number of existing systems, such as Compact Disc (CD), LaserVision (LV) and Digital Optical Recording (DOR). The vertical dashed lines indicate f_c for particular combinations of NA and λ . The two sloping lines represent extrapolations, one based on the CD system, assuming a quadratic relation between f_c and B , the other on a model calculation of the noise-limited situation in which B is proportional to $f_c^{3/2}$.

- [2] Special issue on the Compact Disc, Philips Tech. Rev. 40, 149-180, 1982.
- [3] H. C. Haverkorn van Rijsewijk, P. E. J. Legierse and G. E. Thomas, Manufacture of LaserVision video discs by a photopolymerization process, Philips Tech. Rev. 40, 287-297, 1982; J. G. Kloosterboer, G. J. M. Lippits and H. C. Meinders, Photopolymerizable lacquers for LaserVision video discs, Philips Tech. Rev. 40, 298-309, 1982.
- [4] G. C. Kenney *et al.*, An optical disk replaces 25 mag tapes, IEEE Spectrum 16, No. 2 (February), 33-38, 1979; K. Bulthuis *et al.*, Ten billion bits on a disk, IEEE Spectrum 16, No. 8 (August), 26-33, 1979; L. Vriens and B. A. J. Jacobs, Digital optical recording with tellurium alloys, Philips Tech. Rev. 41, 313-324, 1983/84.
- [5] D. J. Gravesteijn and J. van der Veen, Organic-dye films for optical recording, Philips Tech. Rev. 41, 325-333, 1983/84.
- [6] H. van Tongeren and M. Sens, Write-once phase-change recording in GaSb, Topical Meeting on Optical Data Storage, Technical Digest Series 1987, Vol. 10, (Optical Society of America, Washington D.C., 1987), pp. 50-53; D. J. Gravesteijn, H. M. van Tongeren, M. Sens, T. Bertens and C. J. van der Poel, Phase-change optical data storage in GaSb, Appl. Opt. 26, 4772-4776, 1987.
- [7] M. Hartmann, B. A. J. Jacobs and J. J. M. Braat, Erasable magneto-optical recording, Philips Tech. Rev. 42, 37-47, 1985.
- [8] D. J. Gravesteijn, New material developments for write-once and erasable phase-change optical recording, Topical Meeting on Optical Data Storage, Technical Digest Series 1987, Vol. 10, (Optical Society of America, Washington D.C., 1987), pp. 40-43; also published in Appl. Opt. 27, 736-738, 1988.
- [9] C. M. J. van Uijen, Measured performance of modulation strategies for digital optical recording media, Topical Meeting on Optical Data Storage, Technical Digest Series 1987, Vol. 10, (Optical Society of America, Washington D.C., 1987), pp. 102-105.
- [10] B. H. Verbeek, D. Lenstra and A. J. den Boef, Noise due to optical feedback in semiconductor lasers, Philips Tech. Rev. 43, 292-302, 1987.

use of new materials for semiconductor lasers^[11]. Another way of doing this is to use the method of frequency doubling briefly mentioned earlier. A large world-wide effort is being devoted to frequency-doubled laser systems in which second harmonics are generated in nonlinear materials. The relative ease of producing III-V laser-pumped Nd:YAG lasers, in which frequency-doubling is applied to give a wavelength of 530 nm, has stimulated the search for similar techniques for versions of the 800-nm semiconductor laser with direct-doubling and high power (at 400 nm). In view of the power densities that this would require from the III-V laser, it would be neces-

cal elements with a varying refractive index ('gradient index') has emerged as a viable technology. These new technologies have permitted the design and construction of light paths that are not only less expensive to manufacture, but are also relatively simple and light in weight. More recent developments in the area of light paths incorporating beam splitters based on grating operation will strengthen this trend. Finally, there are also the fascinating developments in integrated planar light paths and the use of holographic focusing techniques and optical-fibre technology. These are the forerunners of optical components that can easily be incorporated in high-performance actuators for use in

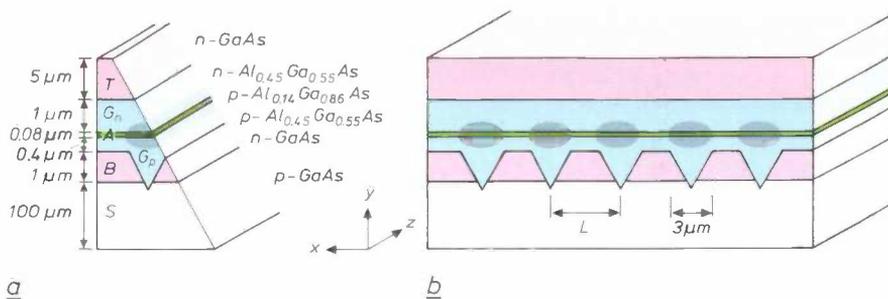


Fig. 4. *a*) Cross-section through the layer structure in a double-heterojunction semiconductor laser. On a substrate *S* of p-GaAs there are, in order, a current-isolating layer *B*, a confinement layer *G_p*, an active layer *A*, in which the laser operation takes place, a confinement layer *G_n* and a top layer *T*, to which contacts can be applied. The V groove etched into the current-isolating layer ensures that the current injected into the structure in the *y*-direction is confined in the *x*-direction. Dimensions and compositions of the various layers are indicated in the figure. The laser light is generated in the grey region and propagates in the *z*-direction. The length of the layer structure in the *z*-direction is much greater than the transverse dimension of the structure. *b*) A number of such lasers placed next to one another form a laser array. If the spacing *L* is large, the lasers can operate independently of one another, and at a spacing of say 150 μm they can in principle be modulated independently, which is useful for parallel recording. If *L* is small (e.g. 5 μm), a composite laser with an increased power is formed because of phase coupling.

sary to consider the application of phase-coupled laser arrays (fig. 4). Success in this area would mean a breakthrough in achievable storage densities, although such a breakthrough would have to be accompanied by a great deal of development activity in recorder optics and on media that can be used in the 400-nm range.

Optics and actuators

The growth of optical recording into the mass market in the consumer area has given an enormous stimulus to the optical industry. It has led to the development of various techniques for the mass production of inexpensive, high-quality optics. These techniques include the hybrid polymer-on-glass replication process and the precision moulding of glass and plastics for the production of aspheric collimator and objective lenses^[12]; see fig. 5. In addition, the use of opti-

cal systems in which the speed of rotation of the disc is increased for high-data-rate applications and in which the random-access facility is greatly improved.

Another challenge is that of the investigation of the strengths and weaknesses of the various modern optical technologies when they are combined with the predicted trends noted earlier towards higher numerical apertures and shorter laser wavelengths.

In the past the servo technology required for dynamic positioning to submicron accuracy has often been called the Achilles heel of optical recording. But now many authorities in this field agree that new servo techniques being developed, new actuator concepts^[13] (see fig. 6) and new electronic possibilities will allow this area to keep pace with developments in the other subsystems. In other words, the control systems for optical recorders should not form a bottleneck for foreseeable developments in high-performance systems.

Recording media

Of the developments in the various subsystems associated with optical-recording technology, those in the recording media are the most diverse. Using the obvious subdivision of replicated, write-once and erasable discs, some general trends can be predicted.

Replicated media^[3] will undoubtedly maintain their current status as carriers of information intended for large-scale distribution. In view of this, the

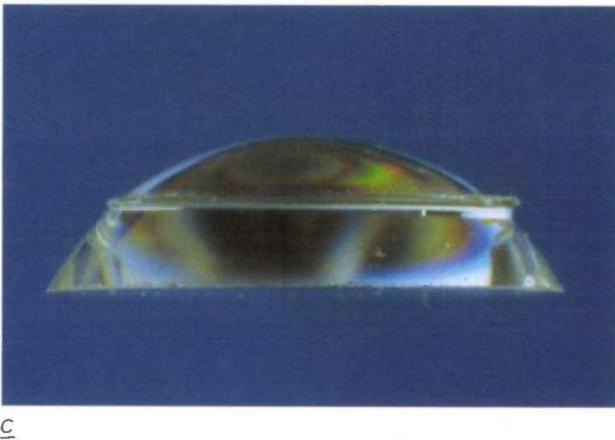
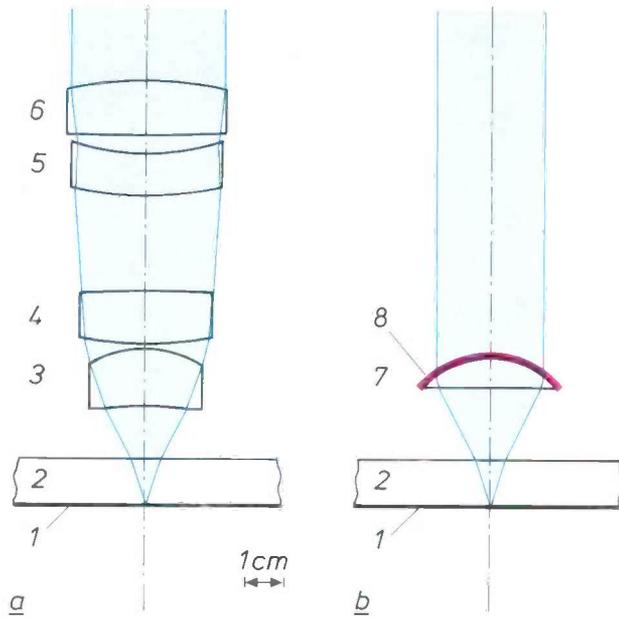


Fig. 5. In optical recording an objective lens focuses the laser beam on to the recording layer (1) on the 'reverse' side of the transparent optical disc (2). If the consequences of the lens aberrations are to be kept within acceptable limits, something other than a single spherical lens must be used. a) One known solution is to use a number of lenses (3 to 6), as in the classical microscope objective. The objective shown here has a numerical aperture of 0.45, and with its mount it weighs about 1.5 grams. b) If an *aspheric* lens is used it is possible to obtain an optically equivalent objective made in one piece and weighing only 20 mg. A lens of this type can be made by the 'polymer-on-glass' replication technology. A thin layer of polymer is applied to a spherical glass preform (7)^[11]. The thickness of the polymer coating (8) is not constant, but is varied in an accurately predetermined way from 0 to 14 μm to produce a high-quality aspheric lens. c) Photograph of an aspheric lens made in this way.

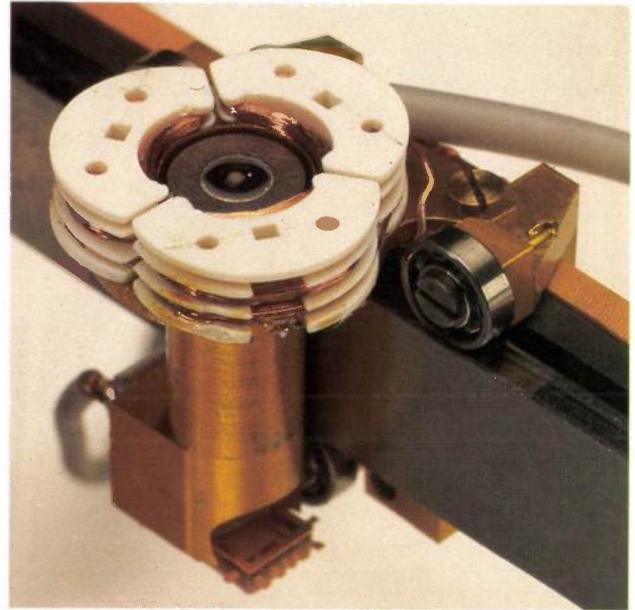


Fig. 6. Like any other rigid body, an objective lens has six degrees of freedom: three orthogonal translations and three rotations. In the actuators now in use only two or three degrees of freedom are normally employed (they are called 2D or 3D actuators). A recent development has led to the 5D actuator shown here, in which the objective lens 'floats' freely surrounded by six separate banana-shaped coils that can control all five of the relevant degrees of freedom (rotation about the axis of symmetry of the lens is irrelevant). The actual lens has a diameter of 3 mm, and is fixed in a permanent-magnet ring with a diameter of 5.5 mm. It can be moved 2 mm in the vertical direction and 1 mm in two perpendicular horizontal directions. The complete actuator is mounted on an arm that can be used to give larger displacements.

necessity to define widely accepted system standards for this area will preclude unlimited diversification in the types of replicated discs. The pattern of development here will tend towards reduction of costs and improvement in quality — where quality means fewer inherent errors on the disc itself. The development of simpler mastering techniques and cost-effective replication technologies for the production of limited series of discs would improve market penetration in the areas of CD-ROM and CD-Audio. The development of new disc formats, which now range from 9 to 30 cm in diameter, will undoubtedly stabilize, since most application areas would now seem to be covered.

The prediction of the future trends in write-once media and systems is uncertain. Although there seems to be a niche in the marketplace for truly permanent

[11] See for example: Semiconductor laser for visible light, Philips Tech. Rev. 44, 23, 1988.

[12] J. Haisma, T. G. Gijsbers, J. J. M. Braat, W. Mesman, J. M. Oomen and J. C. Wijn, Aspherics, Philips Tech. Rev. 41, 285-303, 1983/84.

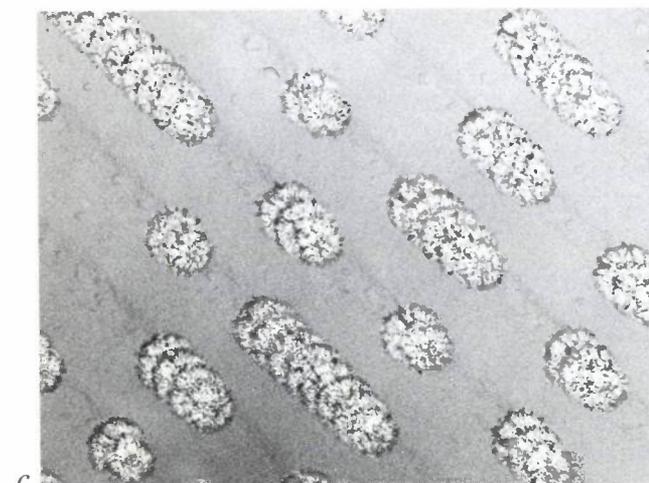
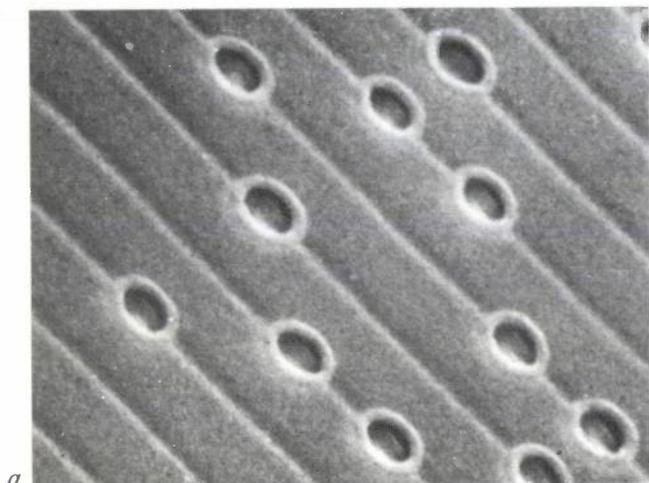
[13] G. E. van Rosmalen, A floating-lens actuator, paper TA7, Int. Symp. on Optical Memory (ISOM '87), Tokyo, Japan, Sept. 1987; also published in Jap. J. Appl. Phys. 26 (suppl. 26-4), 195-197, 1987.

archival information storage — a type of storage for which optical recording technology^[4] offers unique possibilities — the further development of this market will become uncertain once erasable systems are widely available. Nevertheless, it should be mentioned that the signal characteristics of a number of proposed write-once systems — e.g. those using organic-dye

materials^[5] and an amorphous-to-crystalline phase-change^[6] — are so good that they rival replicated discs in quality. Although in the past the properties of some of the ‘ablative’ (hole-forming) write-once systems could deteriorate because of the inherent chemical instability of the materials, a number of the more recently developed systems no longer suffer from this problem, so that truly archival recordings can be made (fig. 7).

The coming years will see the introduction of erasable optical-recording products (fig. 8). It is generally conceded that magneto-optics^[7] is in a very advanced state of development and that this technology will lead the way into the marketplace. This introduction has become a possibility because of the development in recent years of procedures for direct overwriting, of simpler systems for polarization read-out of the signals and of optical-disc technologies that circumvent the problems of the limited chemical stability of the magneto-optical thin films. A great deal of work world-wide is now being concentrated on assessing the limits of magneto-optics in terms of signal characteristics, compatibility with various special modulation methods and useful lifetime.

In addition to magneto-optics, considerable progress has been made in the past year in the development of recording systems in which writing and erasure depend on a reversible amorphous/crystalline phase change^[8]; fig. 8*b*. It is now clear that phase-change recording is feasible — this technology is attractive because of its relative simplicity and reasonable compatibility with the recording and read-out methods for replicated and write-once discs — even though it is not as mature a technology as magneto-optics. A great deal of comparative assessment of the various proposed phase-change systems still has to be done. In some quarters, however, predictions are being made that erasable phase-change systems will one day match the signal quality of known advanced write-once systems and thus become the choice for future generations of high-performance optical recording.



△

Fig. 7. *a*) In ‘ablative’ write-once optical systems a ‘hole’ or ‘pit’ is formed in a recording layer by local heating with a laser beam. This photograph shows an example of digital optical recording in which a tellurium alloy is used as the recording layer. *b*) Example of ablative optical recording in which the recording layer is an organic dye. The photograph relates to analog optical recording, e.g. of video information. *c*) A modern alternative to ablative methods is based on the difference in optical properties between the amorphous and crystalline states of a material such as gallium antimonide. In this method a short pulse from a ‘write’ laser forms small domains with a crystalline structure in an amorphous layer. The result is extremely stable, both chemically and physically. It will therefore have a long life and be very suitable for archival recording. In the example of digital recording shown here the individual crystalline domains are so close together that they overlap. (In this figure and the next one the spacing between adjacent tracks is always about 1.6 μm.)

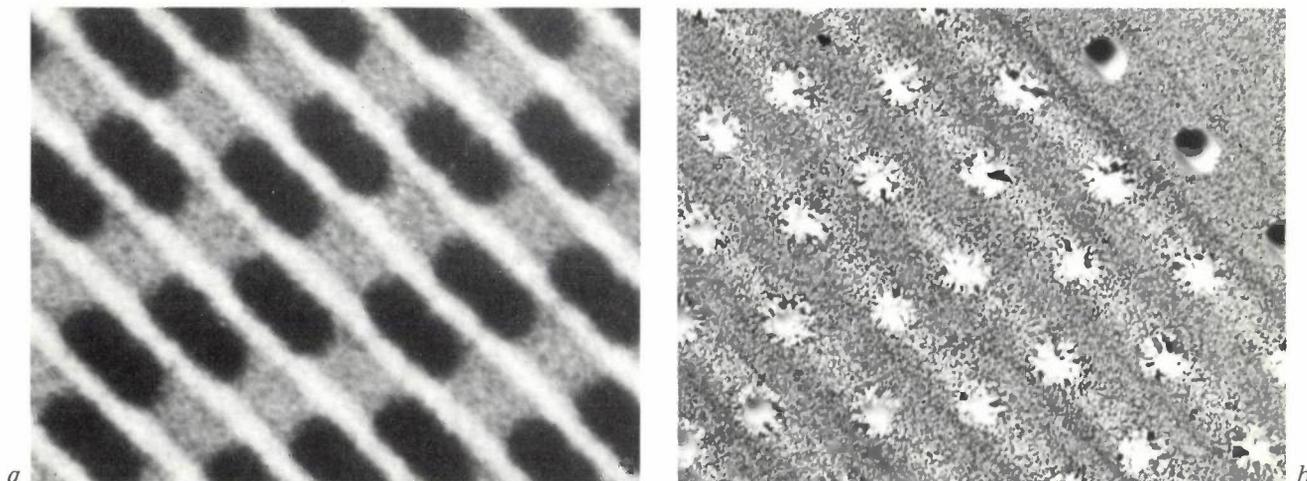


Fig. 8. *a)* Magneto-optic effects can be used for erasable optical recording. A magnetic recording layer is initially magnetized perpendicular to the surface; a laser can then be used in conjunction with a constant external magnetic field to reverse the direction of magnetization locally. Read-out depends on the rotation of the plane of polarization of incident laser light. The information can be erased by irradiating with a laser beam in the presence of a constant external magnetic field, but now with the same direction as the original magnetization. *b)* A more recent development is erasable optical recording based on a reversible amorphous/crystalline phase change. Here the recording layer is of gallium antimonide or indium antimonide, doped with other elements. Initially the recording layer is in the *crystalline* state, and small *amorphous* domains can be created in it by heating rapidly with a laser beam to just above the melting point (the three 'black' areas in the upper right-hand corner of the photograph are amorphous domains). These have different reflection properties from the surrounding crystalline material. The crystalline structure can be restored by heating again with a laser beam to just below the melting point: this erases the stored information. Many erased domains can be identified in the photograph; their reflection properties are again the same as for the surrounding material.

As the demands for extremely high storage densities increase, the search for new optical recording systems that meet the requirements will continue. In addition to the developments mentioned above — which we can assess reasonably well in the light of the knowledge we now have — entirely new methods will undoubtedly emerge. It is almost certain that older ideas such as multilevel recording will be reinvestigated. New techniques such as 'spectral hole burning', in which the optical properties of a recording layer are modified wavelength-selectively and the bit density can increase by several orders of magnitude, offer exciting prospects. In this short summary the emphasis

has been on the evolution of current systems to higher-performance systems. But in a field as young as optical recording we should not discount the probability of a revolution leading to entirely new possibilities.

Summary. The relatively short history of optical recording can be divided into a number of periods. First there is the development of the basic read-only systems that are intended for playing replicated discs, then come the systems and media that can also be used for recording, and finally there is the appearance of erasable optical recording. The future challenge will be to develop systems with higher densities and higher data rates. This article looks at the resultant consequences for the subsystems of optical recording — from laser to disc.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brevannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale, P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

N. J. van Veen	E	Influence of the detector response function in the determination of concentration depth profiles from angle-resolved electron spectroscopy	Anal. Chem. 59	2088-2091	1987
P. Dolizy & F. Groliere	L	Optical method for heterogeneity analysis in thin absorbing films	Appl. Opt. 26	2401-2406	1987
R. L. S. Devine, C. T. Foxon, B. A. Joyce, J. B. Clegg & J. P. Gowers	R	Beryllium diffusion across GaAs/(Al,Ga)As heterojunctions and GaAs/AlAs superlattices during MBE growth	Appl. Phys. A 44	195-200	1987
J. J. Kelly & A. C. Reynders	E	A study of GaAs etching in alkaline H ₂ O ₂ solutions	Appl. Surf. Sci. 29	149-164	1987
S. van Lerberghe	E	Adaptivity in parallel all-pole lattice filters	Digital Signal Processing — 87, V. Cappellini & A. G. Constantinides (eds), Elsevier Science, Amsterdam	125-128	1987
S. Stotz	A	Shift of the morphotropic phase boundary in the PZT system under the influence of electric fields and uniaxial stresses	Ferroelectrics 76	123-132	1987
J. L. W. Kessels	E	PHILAN: A LAN providing a reliable message service	High Speed Local Area Networks, O. Spaniol & A. Danthine (eds), Elsevier Science, Amsterdam	201-212	1987
E. Delhaye, C. Rocher, J.-C. Baelde, J.-M. Gibereau & M. Rocchi	L	A 2.5-ns, 40-mW, 4 × 4 GaAs multiplier in two's complement mode	IEEE J. SC-22	409-414	1987
M. J. M. Pelgrom & H. A. H. Termeer	E	A 32-kbit variable-length shift register for digital audio application	<i>ibid.</i>	415-422	1987
M. J. M. Pelgrom, H. E. J. Wulms, P. van der Stokker & R. A. Bergamaschi	E	FEBRIS: A chip for pattern recognition	<i>ibid.</i>	423-429	1987
P. Sheridan & C. M. Huizer	E	An expression for the propagation delay of a differential split-level (DSL) CMOS logic gate	<i>ibid.</i>	457-459	1987
P. Jean (<i>TRT, Le Plessis Robinson</i>), V. Pauker & P. Dautriche	L	Wide-band monolithic GaAs phase detector for homodyne reception	IEEE MTT-S Int. Microwave Symp. Digest, Las Vegas, NV, 1987	169-171	1987
J. H. M. Neijzen & J. H. A. Glashörster (<i>Philips Centre for Manufacturing Technol., Eindhoven</i>)	E	Fretting corrosion of tin-coated electrical contacts	IEEE Trans. CHMT-10	68-74	1987

- | | | | | | |
|--|----------|---|---------------------------------|-----------|------|
| J. J. Goedbloed | <i>E</i> | Transients in low-voltage supply networks | IEEE Trans. EMC-29 | 104-115 | 1987 |
| K. A. Schouhamer Immink & G. F. M. Beenker | <i>E</i> | Binary transmission codes with higher order spectral zeros at zero frequency | IEEE Trans. IT-33 | 452-454 | 1987 |
| P. N. Kuin | <i>E</i> | The interpretation of rheograms of magnetic lacquers measured with a couette apparatus | IEEE Trans. MAG-23 | 97-99 | 1987 |
| J. P. C. Bernards, C. P. G. Schrauwen, S. B. Luitjens, V. Zieren & R. W. de Bie | <i>E</i> | Material and recording properties of perpendicular CoCr media | <i>ibid.</i> | 125-127 | 1987 |
| J. J. Brondijk, P. E. Wierenga, E. E. Feekes* & W. J. J. M. Sprangers* (*P.D. Magnetics, Oosterhout) | <i>E</i> | Roughness and deformation aspects in calendaring of particulate magnetic tape | <i>ibid.</i> | 146-149 | 1987 |
| M. H. Azarian & P. N. Kuin | <i>E</i> | Determination of the particle size distribution of magnetic pigments | <i>ibid.</i> | 192-194 | 1987 |
| D. Hennings | <i>A</i> | Barium titanate based ceramic materials for dielectric use | Int. J. High Technol. Ceram. 3 | 91-111 | 1987 |
| J. H. Streng | <i>E</i> | Calculation of the surface pressure on a vibrating circular stretched membrane in free space | J. Acoust. Soc. Am. 82 | 679-686 | 1987 |
| P. W. J. M. Boumans & J. J. A. M. Vrakking | <i>E</i> | Inductively coupled plasmas: Line widths and shapes, detection limits and spectral interferences. An integrated picture | J. Anal. At. Spectrom. 2 | 513-525 | 1987 |
| P. Suchet, M. Duseaux, J. Maluenda & G. M. Martin | <i>L</i> | Effects of dislocations on threshold voltage of GaAs field-effect transistors | J. Appl. Phys. 62 | 1097-1101 | 1987 |
| E. P. Honig, P. E. Wierenga & J. H. M. van der Linden | <i>E</i> | Theory of elastic behavior of composite materials | <i>ibid.</i> | 1610-1612 | 1987 |
| O. Boser | <i>N</i> | Statistical theory of hysteresis in ferroelectric materials | <i>ibid.</i> | 1344-1348 | 1987 |
| D. Moroni, J. P. André, E. P. Menu, P. Gentric & J. N. Patillon | <i>L</i> | Photoluminescence investigation of InGaAs-InP quantum wells | <i>ibid.</i> | 2003-2008 | 1987 |
| C. van Opdorp & M. R. Leys | <i>E</i> | On the factors impairing the compositional transition abruptness in heterojunctions grown by vapour-phase epitaxy | J. Cryst. Growth 84 | 271-288 | 1987 |
| P. F. Bordui, J. C. Jacco, G. M. Loiacono, R. A. Stolzenberger & J. J. Zola | <i>N</i> | Growth of large single crystals of KTiOPO_4 (KTP) from high-temperature solution using heat pipe based furnace system | <i>ibid.</i> | 403-408 | 1987 |
| M. Delfino, M. Jaczynski, A. E. Morgan, C. Vorst, M. E. Lunnon & P. Maillot | <i>S</i> | Gettering of copper in silicon-on-insulator structures formed by oxygen ion implantation | J. Electrochem. Soc. 134 | 2027-2030 | 1987 |
| M. F. H. Schuurmans, H. W. A. M. Rompa & R. Eppenga | <i>E</i> | Possibility of green light emission from GaP/AlP (001) superlattices | J. Lumin. 37 | 269-273 | 1987 |
| P. C. M. Gubbens*, A. M. van der Kraan* (*Interuniv. Reactor Inst., Delft), R. P. van Stapele & K. H. J. Buschow | <i>E</i> | ^{169}Tm Mössbauer effect in $\text{Tm}_2\text{Te}_{14}\text{B}$ | J. Magn. & Magn. Mater. 68 | 238-242 | 1987 |
| F. J. J. Blommaert* & J. J. Neve* (*Inst. for Perception Res., Eindhoven) | | Reading fields of magnifying loupes | J. Opt. Soc. Am. A 4 | 1820-1830 | 1987 |
| J. W. C. de Vries | <i>E</i> | Resistivity of thin Au films as a function of grain diameter and temperature | J. Phys. F 17 | 1945-1952 | 1987 |
| M. J. J. Theunissen, H. J. J. Jaspers, J. M. G. Hanssen & M. L. Verheijke | <i>E</i> | Neutron activation analysis of the process of metal gettering in silicon slices | J. Radioanal. & Nucl. Chem. 113 | 391-396 | 1987 |
| M. L. Verheijke, H. J. J. Jaspers, J. M. G. Hanssen & M. J. J. Theunissen | <i>E</i> | Application of neutron activation analysis in the field of silicon technology for the modern electronic industry | <i>ibid.</i> | 397-403 | 1987 |
| P. K. Larsen, G. Meyer-Ehmsen (Univ. Osnabrück), B. Bölger & A.-J. Hoeven | <i>E</i> | Surface disorder induced Kikuchi features in reflection high-energy electron diffraction patterns of static and growing GaAs(001) films | J. Vac. Sci. & Technol. A 5 | 611-614 | 1987 |

- G. G. P. van Gorkom & A. M. E. Hoeberechts *E* Silicon cold cathodes as possible sources in electron lithography systems *J. Vac. Sci. & Technol. A 5* 1544-1548 1987
- D. R. Wolters & A. T. A. Zegers-van Duijnhoven *E* On dielectric breakdown in oxidized silicon *ibid.* 1563-1568 1987
- G. E. Thomas *E* Thin films for optical recording applications *ibid.* 1965-1966 1987
- D. R. Wolters & A. T. A. Zegers-van Duynhoven *E* On the initial regime of silicon oxidation *Phil. Mag. B 55* 669-672 1987
- P. Friedel & J.-P. Landesman *L* Photoemission study of the passivation of GaAs in a nitrogen multipolar plasma *ibid.* 711-719 1987
- J. S. Cohen & A. G. Schlijper (*Shell Lab., Rijswijk*) *E* Long-range order in $Al_xGa_{1-x}As$ *Phys. Rev. B 36* 1526-1530 1987
- G. E. W. Bauer & T. Ando (*Univ. Tokyo*) *E* Comment on 'Excitonic coupling in GaAs/GaAlAs quantum wells in an electric field' *Phys. Rev. Lett. 59* 601 1987
- H. H. A. Smit*, R. C. Thiel* (**Univ. Leiden*) & K. H. J. Buschow *E* ^{155}Gd Mössbauer effect in $Gd_2Co_{14}B$ *Physica 145B* 329-334 1987
- P. Gamand, M. Fairburn, C. Varin & J.-C. Meunier *L,R* 1 to 20 GHz monolithic distributed amplifier using GaAs MESFET's or HEMT's *Proc. 17th Eur. Microwave Conf., Rome 1987* 249-254 1987
- M. Levent-Villegas, M. Soulard & C. Villalon *L* S-band GaAs monolithic linear dual phase modulator *ibid.* 421-426 1987
- R. H. Coursant, P. Eyraud*, L. Eyraud* (**INSA, Villeurbanne*), M. Fink (*Univ. Paris*) & J. M. Tellier *L* Preparation and characterization of lead titanate and lead metaniobate piezoceramics for ultrasonic transducer design *Proc. 6th IEE Int. Symp. on Applications of ferroelectrics, Bethlehem, PA, 1986* 442-447 1986
- R. H. Coursant, C. Méquio & J. M. Tellier *L* Automatic method of intrinsic characterization of lossy piezoelectric structures *ibid.* 730-734 1986
- P. Rabinzohn, J. Bellaiche, C. Rocher & S. Gourrier *L* Reactive ion etching of tungsten nitride in CHF_3-SF_6 based gas mixtures using a statistical design approach *Proc. 8th Int. Symp. on Plasmas chemistry, Vol. 2, Tokyo 1987* 1044-1049 1987
- M. Rocchi *L* GaAs digital IC's for multigigabit's systems *Proc. ESSCIRC '87, Bad Soden 1987* 149-156 1987
- A. J. Guest *R* Modelling microchannel plate (MCP) performance in thin, flat CRTs *Proc. Eurodisplay '87* 208-211 1987
- P. Houdy, V. Bodart, L. Nénot*, D. Corno*, B. Pardo* (**Inst. d'Optique Théorique et Appliquée, Orsay*), M. Arbaoui*, N. Alehyane*, R. Barchewitz* (**Univ. Paris*), Y. Lepêtre° & E. Ziegler° (**Argonne Nat. Lab., Argonne, IL*) *L* Capability of kinetic ellipsometry as an *in situ* control system for ultra thin layers stacks growth: Application to the realization of performant soft X-ray optics *Proc. Spie 688* 122-128 1987
- P. Friedel, J. P. Landesman & R. Mabon *L* Etude par photoémission de la passivation de GaAs en plasmas multipolaires d'azote et d'hydrogène *Rev. Phys. Appl. 22* 797-802 1987
- J. B. Clegg, I. G. Gale, G. Blackmore (*R. S. R. E., Great Malvern*), M. G. Dowsett*, D. S. McPhail* (**City of London Polytech., London*), G. D. T. Spiller (*British Telecom Res., Martlesham Heath*) & D. E. Sykes (*Loughborough Consultants, Loughborough*) *R* A SIMS calibration exercise using multi-element (Cr, Fe and Zn) implanted GaAs *Surf. & Interface Anal. 10* 338-342 1987
- G. Delmas, J.-P. Hazan, M. Steers & C. Piaget *L* La domotique ou une maison au service de l'habitant *Télécom No. 72* 21-22 1987
- P. C. P. Bouten *E* Lifetime of pristine optical fibers *Thesis, Eindhoven* 1-140 1987
- W. J. van Gils *E* Design of error-control coding schemes for three problems of noisy information transmission, storage and processing *ibid.* 1-231 1988
- P. L. Holster & J. A. H. Jacobs *E* Theoretical analysis and experimental verification on the static properties of externally pressurized air-bearing pads with load compensation *Tribology Int. 20* 276-289 1987



J. J. J. Bastiaens and W. C. H. Gubbels, The 256-kbit SRAM: an important step on the way to submicron IC technology, Philips Tech. Rev. 44, No. 2, 33-42, April 1988.

The static random-access memory (SRAM) with a capacity of 256 kbit, developed at Philips Research Laboratories, contains details of 1.2 μm . The cells in this memory, made in full CMOS technology (complementary metal-oxide semiconductor), consist of four transistors, plus two connecting transistors. The current requirement and access time of the SRAM are both small, because of the application of the 'dynamic double-word-line technique'. Scaling down the details in an integrated circuit entails various problems. Shortening the transistor channel causes the current leakage known as 'punch-through' and premature failure of the transistor due to 'hot' electrons. Narrowing the transistor channel leads to difficulties because of the 'bird's-beak' effect, which raises the threshold voltage. These problems have been solved by adapting the technology, including the change to LDD transistors (LDD: lightly doped drain). Other problems can arise from contacts between the interconnection layers and the resistance of the conductors.

G. E. Thomas, Future trends in optical recording, Philips Tech. Rev. 44, No. 2, 51-57, April 1988.

The relatively short history of optical recording can be divided into a number of periods. First there is the development of the basic read-only systems that are intended for playing replicated discs, then come the systems and media that can also be used for recording, and finally there is the appearance of erasable optical recording. The future challenge will be to develop systems with higher densities and higher data rates. This article looks at the resultant consequences for the subsystems of optical recording — from laser to disc.

T. Blaffert, EXPERTISE: an expert system for infrared spectrum evaluation, Philips Tech. Rev. 44, No. 2, 44-50, April 1988.

Infrared spectra are exceptionally useful for providing information about chemical structure. Analytical chemists who use these spectra to identify the chemical structure of unknown or newly synthesized compounds will find the EXPERTISE expert system a great help. Besides identifying complete spectra by comparing them with the spectra in an extensive library, this system can to some extent 'reason' as the chemist does. This is achieved by subdividing the structures in the library into substructures, establishing the spectral features related to these substructures, recognizing these features in spectra of unidentified compounds and combining the substructures to form a complete structure. The procedure for identifying the substructures from the spectral features includes a safety factor to prevent substructures from being excluded in error because of statistical variations in the spectra. This is done with the aid of fuzzy-set theory. The system may provide more than one result, in which case the chemist has to make the final decision. The great advantage of such systems is that all the possible results are found relatively quickly.

I wish to subscribe to

PHILIPS TECHNICAL REVIEW

(date)

(signature)

Please tick the appropriate box

Regular subscription 80 guilders or U.S. \$ 35.00 per volume

The subscription includes postage and will start with Vol. 44, No. 1.
Please pay when you receive our invoice.

Student's subscription 32 guilders or U.S. \$ 14.00 per volume

Please send a copy of your student's card or other written proof that you are a student; valid for two volumes.

Name

Initials

Title

Address

stamp

as

postcard

**Administration Department
Philips Technical Review**

**Philips Research Laboratories
Building WY 136**

P.O. Box 80 000

**5600 JA Eindhoven
The Netherlands**

Ned. Philips Bedrijven B.V.
PHILIPS RESEARCH LABS.
LIBRARY WY - 1
P.O. Box 80.000
5600 JA EINDHOVEN
THE NETHERLANDS

Contents

Wet-chemical etching of III-V semiconductors	61
J. J. Kelly, J. E. A. M. van den Meerakker, P. H. L. Notten and R. P. Tijburg <i>Removing material by wet-chemical etching is an important process in the fabrication of III-V semiconductor components</i>	
Then and Now (1938-1988)	75
Bistability in quantum-well lasers	76
A. I. Kucharska, P. Blood and E. D. Fletcher <i>Special semiconductor-laser structures allow rapid switching between high and low light-output levels</i>	
Chemical modification of surfaces	81
J. J. Ponjé and P. N. T. van Velzen <i>Surface modification — an effective new chemical technology</i>	
Striations in a gas discharge	89
F. C. van den Heuvel <i>The electrical circuit has an effect on optical instabilities in a gas discharge</i>	
Scientific publications	96

Wet-chemical etching of III-V semiconductors

J. J. Kelly, J. E. A. M. van den Meerakker, P. H. L. Notten
and R. P. Tijburg

In recent years the electronics industry has shown growing interest in semiconductors consisting of elements from groups III and V of the Periodic Table. This is largely because multilayer structures of such semiconductors have interesting applications. A well-known example is the semiconductor laser in the Compact Disc player, an important product for consumer electronics. Device structures with the required patterns can be obtained by dissolving parts of the semiconductor materials by a wet-chemical etching technique. Until recently wet etching was a relatively empirical process. An extensive study at Philips Research Laboratories in Eindhoven has provided a much better understanding of the etching behaviour of III-V semiconductors. One result is that the methods and etchants used in present applications can now be optimized more effectively and appropriate etchants can be developed for new applications.

Introduction

For many years the most widely used basic material in the semiconductor industry has been silicon. Since the seventies, however, we have seen the growing emergence of III-V semiconductors such as GaP, GaAs and InP for applications including light-emitting diodes, microwave field-effect transistors^[1] and semiconductor lasers^[2]. For these applications multilayer structures of different III-V semiconductors are used, which are grown epitaxially on a III-V substrate. A well-known combination is that of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ in lasers with an emission wavelength of about 0.8 μm for Compact Disc players and digital optical recording^[3]. Another well-known combination is that of InP and $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ in lasers emitting at wavelengths of 1.30 and 1.55 μm for fibre-optic communications^[4].

Multilayer structures are made with extremely accurate patterns to give them the required characteristics. Besides the mastery of the technique of producing

different III-V semiconductors by epitaxial growth, this also requires an accurate etching technique for removing material at the right places. This is usually done by wet-chemical etching, although dry etching methods like plasma etching are occasionally used. III-V semiconductors can be etched to the required shapes by wet-chemical methods because the materials are generally perfect single crystals and the etch rate depends on the crystal orientation. The structure may often consist of a combination of materials

[1] P. Baudet, M. Binet and D. Boccon-Gibod, Low-noise microwave GaAs field-effect transistor, *Philips Tech. Rev.* **39**, 269-276, 1980.

[2] G. A. Acket, J. J. Daniele, W. Nijman, R. P. Tijburg and P. J. de Waard, Semiconductor lasers for optical communication, *Philips Tech. Rev.* **36**, 190-200, 1976; J. C. J. Finck, H. J. M. van der Laak and J. T. Schrama, A semiconductor laser for information read-out, *Philips Tech. Rev.* **39**, 37-47, 1980.

[3] Special issue 'Compact Disc Digital Audio', *Philips Tech. Rev.* **40**, 149-180, 1982;

K. Bulthuis, M. G. Carasso, J. P. J. Heemskerk, P. J. Kivits, W. J. Kleuters and P. Zalm, Ten billion bits on a disk, *IEEE Spectrum* **16**, No. 8 (August), 26-33, 1979.

[4] G. A. Acket, Halfgeleiderlasers voor optische communicatie, *Ned. T. Natuurk. B* **51**, 13, 1985;

G. A. Acket and W. Nijman, Recente ontwikkelingen op het gebied van halfgeleiders, *Ned. T. Natuurk. A* **53**, 22-24, 1987.

Prof. Dr J. J. Kelly, Professor of Applied Electrochemistry at the State University of Utrecht, was until recently with Philips Research Laboratories, Eindhoven; J. E. A. M. van den Meerakker, P. H. L. Notten and R. P. Tijburg are with Philips Research Laboratories.

which, though chemically closely related, are still sufficiently different for a suitable etchant to give a completely different etching behaviour.

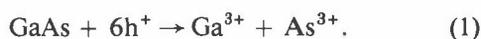
In the investigation described in this article we made an extensive study of the etching behaviour of various III-V semiconductors in a variety of aqueous solutions. This was done for each of the three possible etching mechanisms: electrochemical oxidation with an external voltage source ('anodic etching'), electrochemical etching with an oxidizing agent but no external voltage source, known as 'electroless etching', and chemical etching with a reactive compound. The investigation has led to an improved understanding of the processes involved in the different mechanisms. As a result, some new etchants have been developed that can be used to improve the fabrication of the structures for certain applications.

In this article we shall first give a brief description of the three etching mechanisms. We shall then look at some kinetic and crystallographic aspects of etching that are relevant to the production of particular etch patterns. Finally we shall give some examples of wet-chemical etching of III-V semiconductors in practical applications.

Etching mechanisms

Anodic etching

For anodic etching of a III-V semiconductor in an electrolyte the semiconductor is connected to the positive terminal of a direct-voltage source and an inert counter-electrode is connected to the negative terminal. When the potential of the anode is made sufficiently positive a current starts to flow, and the constituent elements become oxidized at the surface of the semiconductor. The ions formed in this process dissolve. The occurrence of oxidation implies that electrons are withdrawn from the valence band of the semiconductor or — and this amounts to the same thing — that holes are supplied to the valence band. In general, the dissolution of a III-V semiconductor (e.g. GaAs) is found to require six holes (h^+) per formula unit^[5]:



The nature of the reaction products depends on the pH-value and composition of the electrolyte. The etching reaction of GaAs in acidic solutions can be represented as follows:



The dependence of the anodic etch rate on the applied potential is determined by the potential distribution at the interface between the semiconductor

and the electrolyte. As with a metal electrode, there is a potential difference across a region known as the Helmholtz double layer in the solution near the interface. There is an important difference from a metal electrode, however; the potential drop extends some way into the semiconductor material; see *fig. 1*. Because of this 'space-charge layer', which is formed below the surface as a result of the relatively low concentration of charge carriers in a semiconductor, the energy bands are curved.

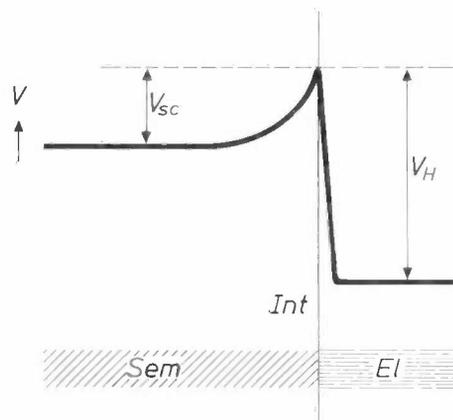


Fig. 1. The potential V at the interface Int of a semiconductor Sem and an electrolyte El . The potential distribution at a semiconductor electrode in an electrolyte differs from that of a metal electrode in that there is not only a potential difference, V_H , across the Helmholtz double layer but also a much more gradual fall in potential, V_{sc} , across the space-charge layer beneath the semiconductor surface.

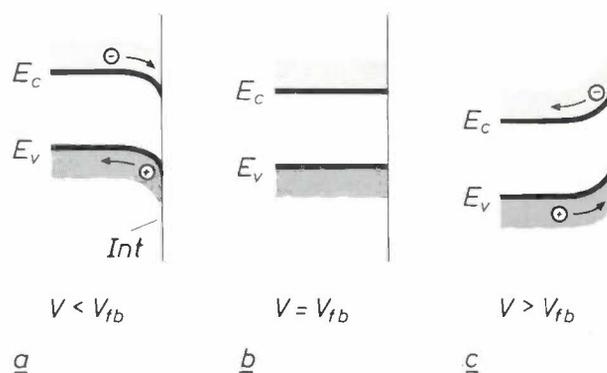


Fig. 2. Energy-level diagram for a semiconductor at the interface Int with an electrolyte for different values of the external potential V with respect to the flat-band potential V_{fb} , i.e. the potential at which the energy-band edges are flat right up to the surface. E_v top of the valence band. E_c bottom of the conduction band. The relative movements of electrons in the conduction band and holes in the valence band are also shown. At $V < V_{fb}$ (a) the bands are bent downwards; electrons tend to move to the surface and holes tend to move towards the bulk of the semiconductor. At $V = V_{fb}$ (b) the concentration of charge carriers at the surface is equal to the concentration in the bulk. At $V > V_{fb}$ (c) the bands are bent upwards, so that holes tend to go to the surface and electrons to the bulk. The location of E_v and E_c at the interface does not depend on the band-bending. The Fermi level (not shown) is assumed to be straight; in the bulk it is close to the conduction band if the semiconductor is n-doped and close to the valence band if the semiconductor is p-doped.

When the external potential is varied it is often found that only the potential difference across the space-charge layer changes, while that across the Helmholtz layer remains unchanged^[6]. There is a particular potential at which the energy-band edges in the semiconductor go straight to the surface without bending. This potential, known as the 'flat-band potential', V_{fb} , depends not only on the semiconductor material but also on the doping and the nature of the electrolyte.

The effect of the applied potential on the energy bands is shown schematically in *fig. 2*. When the potential is negative with respect to V_{fb} the bands bend downwards. This means that there are more electrons and fewer holes near the surface than in the bulk of the material. At a potential equal to V_{fb} the bands run completely straight: the charge-carrier concentration is the same at the surface as in the bulk. When the potential is positive with respect to V_{fb} , the bands bend upwards: there are then fewer electrons at the surface and more holes than in the bulk.

With n-type doping the value of V_{fb} is lower than with p-type doping. This is because V_{fb} is a measure of the position of the Fermi level in the flat-band situation, and because this level lies close to the conduction band for n-type doping and close to the valence band for p-type doping. The difference in V_{fb} is therefore approximately equal to E_g/e , where E_g is the band gap of the semiconductor and e the electronic charge.

For the oxidation of a III-V semiconductor to take place via the reaction with holes at the surface, there must be a sufficient number of holes in the valence band. This will be the case if the p-type doping is high enough as a result of the incorporation of a sufficient number of acceptor atoms. When there is equilibrium between the charge carriers at the surface and in the bulk the concentration of holes at the surface (p_s) is given by the Boltzmann equation:

$$p_s = p_b \exp(eV_{sc}/kT), \quad (3)$$

where p_b is the concentration in the bulk, V_{sc} the potential drop across the space-charge layer in the semiconductor, k is the Boltzmann constant and T the absolute temperature. Assuming that the potential difference across the Helmholtz layer is constant, it follows from the definition of the flat-band potential that:

$$V_{sc} = V - V_{fb}, \quad (4)$$

where V is the applied potential. It will be clear that the dissolution of a p-doped semiconductor requires a potential approximately equal to V_{fb} (*fig. 2b*) or, better still, positive with respect to V_{fb} (*fig. 2c*).

A measured current-potential curve for p-GaAs is shown in *fig. 3a*. As would be expected, the current and hence the etch rate in the vicinity of V_{fb} increases exponentially with the potential, as a result of the exponential increase in the concentration of holes at the surface; see eq. (3).

Undoped or n-doped semiconductors cannot be directly dissolved anodically, because of the absence of holes in the valence band. In this case, however, we can make use of the effect of illumination on semiconductors. Photons of sufficient energy can excite electrons from the valence band into the conduction band, so that holes are formed in the valence band. If the band-bending within the semiconductor is appropriate, the holes are able to reach the surface and the etching reaction can take place; this is referred to as photo-anodic etching.

Fig. 3b shows a current-potential curve for an illuminated n-GaAs electrode in an acidic solution. It

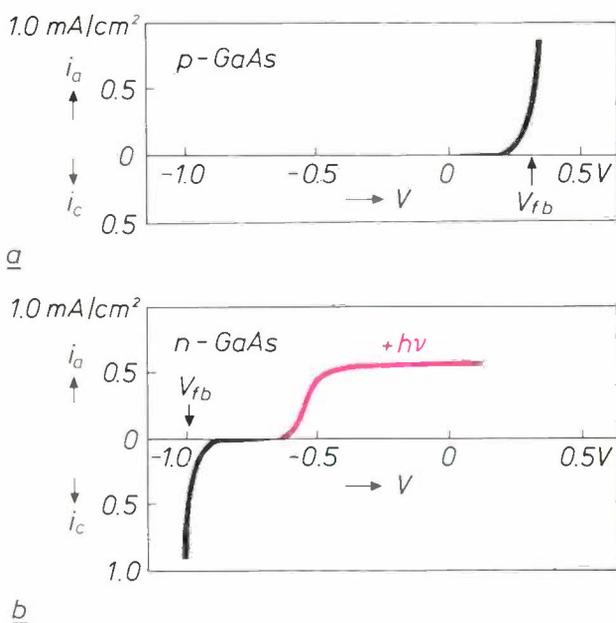


Fig. 3. Current-potential curves for p-GaAs (*a*) and n-GaAs (*b*) in 1N H_2SO_4 . The anodic current density i_a and the cathodic current density i_c are plotted against the potential V of the semiconductor electrode with respect to a saturated calomel electrode. With p-GaAs a large potential-dependent anodic current flows as soon as the potential approaches the flat-band value V_{fb} . With n-GaAs a cathodic current flows at low potentials because electrons from the conduction band reduce H^+ ions in the solution to hydrogen gas. In the dark n-GaAs gives no anodic current. An anodic current (red) does flow upon illumination with photons of sufficient energy $h\nu$. This current only starts to flow at a potential that is substantially higher than the flat-band value for n-GaAs, and does not increase further at high potentials.

^[6] H. Gerischer, Über den Mechanismus der anodischen Auflösung von Galliumarsenid, Ber. Bunsenges. Phys. Chem. **69**, 578-583, 1965.

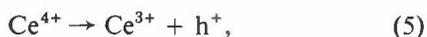
^[6] R. Memming, Energy production by photoelectrochemical processes, Philips Tech. Rev. **38**, 160-177, 1978/79.

can be seen that the anodic current associated with the etching only occurs at a potential much higher than the V_{fb} value of n-GaAs. This is because the holes in an n-type semiconductor are the minority carriers. At a potential close to V_{fb} , the electron concentration at the surface is high and the holes recombine with electrons. At a higher potential the electrons and holes are effectively separated by the electric field (fig. 2c); the holes are now able to reach the surface and cause dissolution of the semiconductor. At a particular potential the anodic current saturates. In this case all the photo-generated holes take part in the etching process and the anodic current is proportional to the intensity of the incident light.

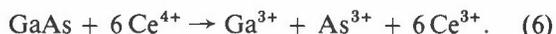
Anodic etching is not normally the most popular method for semiconductor technology. Working with a voltage source and a counter-electrode, with various electrical contacts, can be very inconvenient. Applying an electrical contact to a semiconductor material during the fabrication process can also be a problem and may even be undesirable for certain applications. There are however circumstances in which other etching methods give poor results and anodic etching may offer success. The study of anodic etching can also supply useful information about electroless etching.

Electroless etching

In electrochemical etching without an external voltage source, 'electroless etching', the etchant contains an oxidizing agent. The dissolution of the semiconductor is due to the oxidizing agent depleting the semiconductor of valence-band electrons and thus in fact supplying it with holes, so that it is itself reduced. A suitable oxidizing agent, e.g. for GaAs, is Ce^{4+} :



where the total reaction can be represented by



A reaction of this kind is thermodynamically possible only if the redox potential, i.e. the equilibrium potential of the redox couple (in this case Ce^{4+}/Ce^{3+}), is higher than the potential of the solid (GaAs) in equilibrium with 'its ions' (Ga^{3+} , As^{3+}) in the solution. The rate of the reaction depends on the location of the energy bands of the semiconductor in relation to the energy levels of the redox couple in the solution.

Usually, the redox potential is related to the potential of a reference electrode, e.g. a saturated calomel electrode or a standard hydrogen electrode. In solid-state physics, however, the potential of an electron in vacuum is taken as the reference. If we take the same reference for an electrolyte, then the redox potential corresponds to the energy necessary for transferring

an electron from the redox couple to vacuum. The energy of the electrons in the solution is then defined in the same way as for those in the solid, so that the redox potential can be considered as the Fermi level of the solution. Because of the interaction of the ions of the redox couple with the solvent, the electron energy of the reduced component is not the same as that of the oxidized component. The energy levels of the oxidized and reduced components are not discrete; they are broadened by the fluctuations of the solvation shell. This results in two Gaussian energy-distribution functions that are symmetrical with respect to the Fermi level, i.e. the redox potential; see fig. 4.

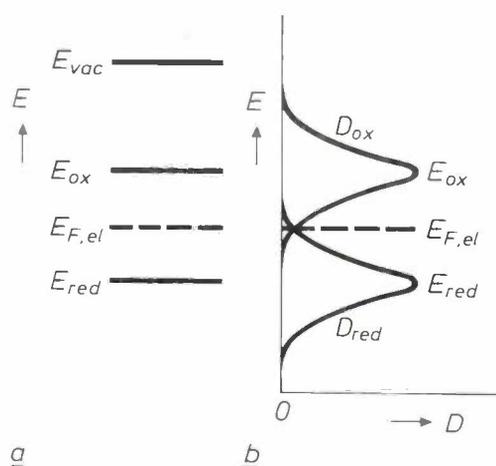


Fig. 4. *a*) Simplified diagram of the energy levels of the components of a redox couple in an electrolyte with respect to the potential energy E_{vac} of an electron in vacuum^[6]. Because of the interaction with the solvent, the empty level E_{ox} of the oxidized component is higher than the filled level E_{red} of the reduced component. The Fermi level $E_{F,el}$, corresponding to the redox potential, lies halfway between E_{red} and E_{ox} . *b*) In reality the levels are broadened to energy-distribution functions D_{red} and D_{ox} owing to fluctuations of the solvation shell.

The dissolution reaction of eq. (6), where the redox couple causes the transfer of six holes to the valence band of the III-V semiconductor, is only possible if the energy-distribution function of the oxidized component corresponds approximately to the location of the valence band of the semiconductor. By measuring the electrode impedance, the potentials of the valence-band and conduction-band edges of the semiconductor can be determined with respect to the same reference potential in the solution. In fig. 5 redox potentials of a number of redox couples in an acidic solution ($pH = 0$) are compared with the band-edge potentials of three widely used semiconductors: GaP, GaAs and InP. The couple Ce^{4+}/Ce^{3+} has a highly positive redox potential, so that holes can be injected into all three semiconductors. The redox potential of the couple Fe^{3+}/Fe^{2+} is less positive; this couple can

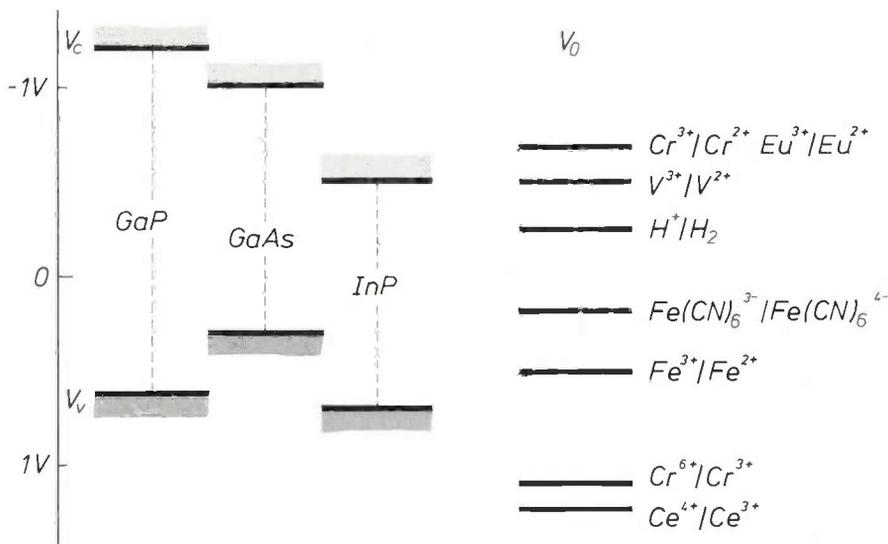


Fig. 5. Potentials of the energy bands of three semiconductors and the redox potentials of a number of redox couples in solutions with a pH of 0, measured with respect to a saturated calomel electrode. The potentials of the band edges of the semiconductors are indicated by V_v and V_c . The standard redox potential V_0 is shown for the redox couples; this is the redox potential measured when the concentration of the oxidized component in the solution is equal to that of the reduced component.

still inject holes into GaAs, but not into GaP and InP. The redox potential of the couple $Fe(CN)_6^{3-}/Fe(CN)_6^{4-}$ is such that the injection of holes is no longer possible, not even into GaAs.

Oxidation with a redox couple can be thought of as the sum of an anodic process and a cathodic process, each with its own partial current-potential curve. Fig. 6 shows these curves for the oxidation of p-GaAs with a Ce^{4+} solution. The anodic partial current increases steeply with the potential, but the cathodic partial current is practically independent of the potential in a wide range. This is because the rate at which the Ce^{4+} ions are reduced is determined entirely by the rate at which they diffuse to the semiconductor surface. Since there is no external current in electroless etching, the etching reaction takes place at the 'mixed potential', where the anodic and cathodic partial currents are equal. The value of the partial currents at the mixed potential determines the etch rate. In this case the etch rate is therefore completely controlled by the diffusion rate of the Ce^{4+} ions in the solution.

Since electroless etching depends on the transfer of electrons from the valence band to the oxidizing agent in the solution, both p-doped and n-doped semiconductors can be dissolved with the same etchants. The current-potential curves differ, however, because the injected holes in n-doped material are minority carriers. As an example, fig. 7 gives the current-potential curves for n-GaAs in a Ce^{4+} solution. In the range of potentials from -1.0 V to -0.6 V the total current is still equal to the cathodic partial current because of

the Ce^{4+} reduction. At these potentials there is not much band-bending, so that the electron concentration is relatively high at the surface and the injected holes recombine with the electrons supplied from the bulk. At higher potentials the electron concentration at the surface becomes negligible and the injected holes can no longer recombine with electrons. The holes then remain at the surface and oxidize the GaAs. Over the entire range of potentials the Ce^{4+} reduction gives the same cathodic partial current, which depends on the diffusion rate of the Ce^{4+} ions. Since at high potentials all the injected holes are used for oxidation, the anodic partial current at such potentials is equal to the cathodic partial current. At the mixed potential the rates of the GaAs oxidation and Ce^{4+} reduction are exactly equal and the recombination rate is zero. The etch rate of n-GaAs is thus determined by the diffusion rate of the Ce^{4+} ions and is the same as that of p-GaAs.

In fig. 5 the band-edge potentials of the semiconductors and the redox potentials of the redox couples apply to acidic solutions ($pH = 0$). A change in the pH -value of the solution has no effect on the redox potential of a couple like $Fe(CN)_6^{3-}/Fe(CN)_6^{4-}$. The presence of H^+ ions or OH^- ions does however have a marked effect on the number of ions adsorbed at the semiconductor surface. This changes the potential difference across the Helmholtz layer (fig. 1), resulting in a different flat-band potential. It is known that the flat-band potential decreases by about 60 mV per unit increment in the pH . This means that at a pH of 14

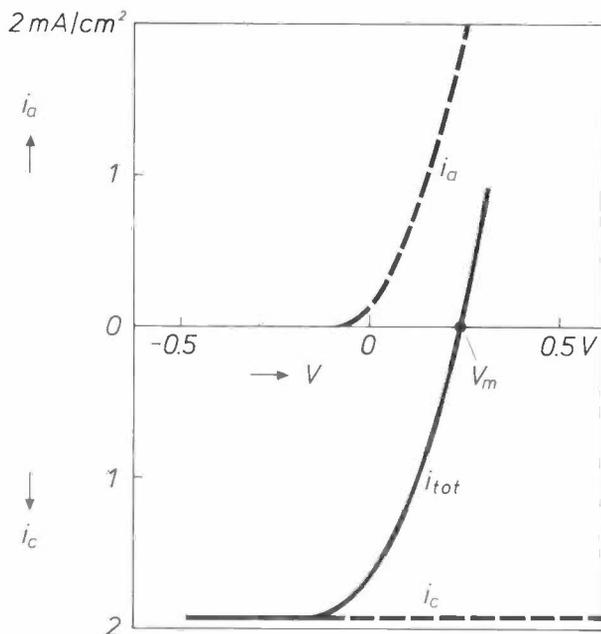


Fig. 6. Current-potential curves for p-GaAs in a Ce^{4+} solution. The anodic partial current density i_a due to the GaAs oxidation, the cathodic partial current density i_c due to the Ce^{4+} reduction, and the total current density i_{tot} are shown, plotted against the potential V with respect to a saturated calomel electrode. Whereas i_a at a particular potential increases steeply with V , the value of i_c is practically independent of the potential in a wide range. The electroless etching takes place at the mixed potential V_m , where $i_a = i_c$. In this case the etch rate, which is proportional to the value of i_a at V_m , is determined entirely by the constant cathodic partial current.

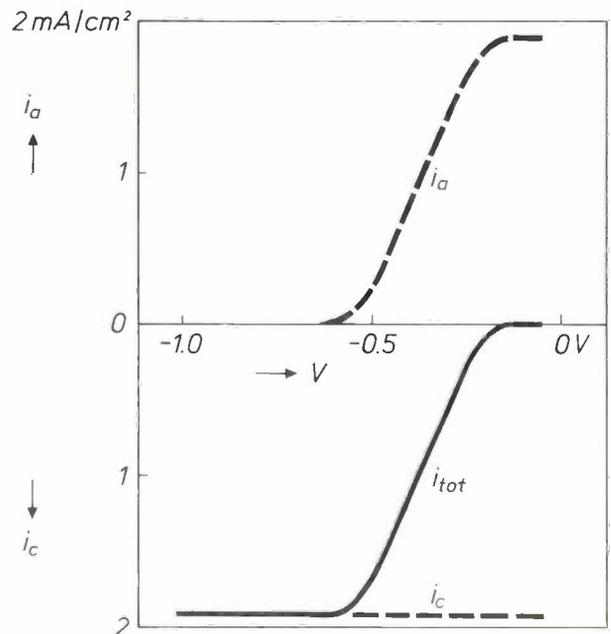


Fig. 7. Current-potential curves (as in fig. 6) for n-GaAs in a Ce^{4+} solution. Here again i_c is constant but i_a also becomes constant at high potentials and practically equal to i_c . The value of i_c again determines the etch rate.

the band-edge potentials of the semiconductors are shifted upwards by about 0.8 V with respect to those in fig. 5. The redox potential of $\text{Fe}(\text{CN})_6^{3-}/\text{Fe}(\text{CN})_6^{4-}$, for example, then comes to lie below the valence-band-edge potentials of GaP, GaAs and InP. The result is that this couple can inject holes into the valence band of all three semiconductors, so that etching is in fact possible at a pH of 14.

A redox potential that is well placed with respect to the valence-band-edge potential of the semiconductor is not always a guarantee of a satisfactory etching reaction. At the surface of InP, for instance, the reduction of Ce^{4+} is inhibited to such an extent, probably due to the formation of a thin oxide film, that there is hardly any etching. Etching GaAs with CrO_3/HF solutions is also less effective. In spite of the appropriate redox potential of the $\text{Cr}^{6+}/\text{Cr}^{3+}$ couple, the dissolution rate of the GaAs is much lower than would be expected from the diffusion rate of Cr^{6+} . In this case the semiconductor surface is partly passivated by the adsorption of an intermediate formed during the etching reaction, probably a complex of hexavalent and trivalent chromium^[7].

Besides the oxidizing agents that can give an etching reaction in the dark, there are other agents that can

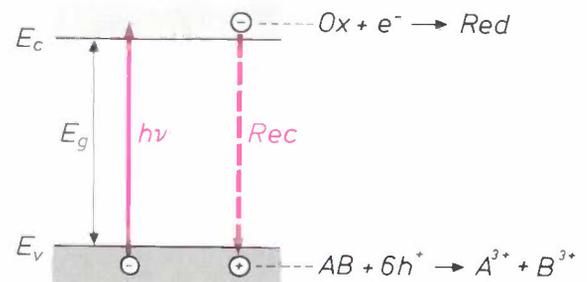


Fig. 8. Principle of electroless photo-etching of a III-V semiconductor AB. Photons of energy $h\nu$ equal to or greater than the band gap E_g excite electrons from the valence band to the conduction band. The holes formed oxidize the semiconductor while the electrons in the conduction band reduce the oxidizing agent Ox to Red. A competing process is the recombination Rec of electrons and holes.

etch III-V semiconductors under illumination. This 'photo-etching' depends on both majority and minority carriers; see fig. 8. Photons with an energy equal to or greater than the band gap of the semiconductor generate electron-hole pairs. The holes that are formed oxidize the semiconductor, while the electrons reduce the oxidizing agent. A competing process is the recombination of electrons and holes. The couples that would appear to be useful for the photo-etching are those whose redox potential lies above the valence-band potential of the semiconductor and whose oxidized component has an energy-distribution function that corresponds to the conduction band. In the case of GaAs, for example, these are the couples $\text{V}^{3+}/\text{V}^{2+}$,

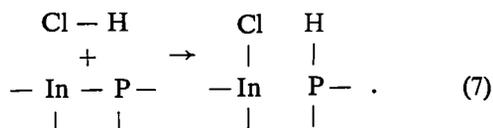
$\text{Cr}^{3+}/\text{Cr}^{2+}$ and $\text{Eu}^{3+}/\text{Eu}^{2+}$ (fig. 5). We have found however that these couples are unsuitable for photo-etching III-V semiconductors, because the electron transfer to the oxidizing agent cannot compete with the electron-hole recombination. There are some more complicated redox couples that do provide effective electron transfer [8], however. Acidic H_2O_2 solutions, for example, can be used to photo-etch both p-type and n-type GaAs [9]. We shall not consider the intricate mechanism of these etching reactions here.

Chemical etching

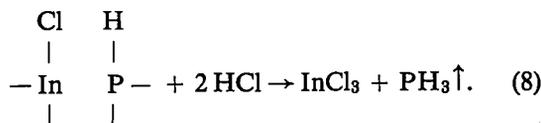
Chemical etching does not involve free charge carriers and cannot therefore be affected by an external potential. In chemical etching the surface of the semiconductor is attacked through the action of reactive molecules (e.g. HCl), with the result that bonds between the surface atoms are broken and at the same time new bonds are formed with the atoms of the reactive molecules. The compounds formed dissolve in the etchant or escape as gases.

A mechanism of this type is found in the etching of InP with concentrated HCl solutions [10]. Since the dissolution of InP is due to the reaction with undissociated HCl molecules, the HCl concentration has to be fairly high — higher than 6 M at room temperature. The concentration can be lower in a solution containing a large amount of acetic acid, because the dissociation content of HCl is smaller in such a solution [11].

In the reaction between InP and HCl it is likely that a bond between In and P is broken at the same time as the bond between H and Cl, and that new bonds are then formed between In and Cl and between P and H:



Since each surface atom is bonded to three atoms below it, two bonds then remain to be broken. Finally, the indium dissolves as InCl_3 , while the phosphorus disappears as PH_3 :



In some cases the etching mechanism may be changed by the action of light. An example is found in the etching of GaAs with H_2O_2 solutions. In the dark only a chemical mechanism is active, but illumination gives rise to an electrochemical etching process, in which the holes formed in the valence band oxidize the GaAs [8].

Kinetic and crystallographic aspects

The rate at which III-V semiconductors are dissolved by a particular etching mechanism depends on a number of factors. The rate, for example, may be determined entirely by the kinetics of the reaction at the semiconductor surface. However, if the rate of this reaction is very high, the rate-determining parameter may be mass transfer in the solution. The etch rate may also be very dependent on the crystallographic orientation of the surface. Etching near the edges of masks may produce special effects. If a semiconductor is being etched in contact with another semiconductor or a metal, it might be expected that the rate of dissolution would be affected by the galvanic interaction between the materials. Since all these factors are relevant to the practical applications to be discussed, we shall first take a closer look at them.

Rate-determining step

In a kinetically controlled etching process, where the surface reaction is rate-determining, the concentration of the rate-determining species at the surface (c_s) will be equal to that in the bulk of the solution (c_b). In the simplest approximation the etch rate r_k is then given by

$$r_k = k c_b^n, \quad (9)$$

where k is the reaction-rate constant and n the order of the reaction. The rate constant is given by:

$$k = A \exp(-E_a/RT), \quad (10)$$

where A is a constant, E_a the activation energy for the reaction, R the gas constant and T the absolute temperature. A plot of $\log r_k$ against $1/T$ gives a straight line. The activation energy that can be derived from the slope of this line is of the order of 40 kJ/mol for most kinetically controlled etching reactions. This means that the activation energy per molecule is of the order of 0.4 eV.

[7] J. van de Ven, J. E. A. M. van den Meerakker and J. J. Kelly, The mechanism of GaAs etching in CrO_3 -HF solutions, I. Experimental results, *J. Electrochem. Soc.* 132, 3020-3026, 1985;

J. J. Kelly, J. van de Ven and J. E. A. M. van den Meerakker, The mechanism of GaAs etching in CrO_3 -HF solutions, II. Model and discussion, *J. Electrochem. Soc.* 132, 3026-3033, 1985.

[8] J. J. Kelly, J. E. A. M. van den Meerakker and P. H. L. Notten, Electrochemistry of photo-etching and defect-revealing in III-V materials, in: *Grundlagen von Elektrodenreaktionen*, Dechema Monographien, Band 102, VCH Verlagsgesellschaft, Frankfurt am Main 1986, pp. 453-464.

[9] D. V. Podlesnik, H. H. Gilgen and R. M. Osgood, Jr, Deep-ultraviolet induced wet etching of GaAs, *Appl. Phys. Lett.* 45, 563-565, 1984.

[10] P. H. L. Notten, The etching of InP in HCl solutions: a chemical mechanism, *J. Electrochem. Soc.* 131, 2641-2644, 1984.

[11] W. Huber, Titrations in non aqueous solvents, Academic Press, New York 1967, p. 215 and 226.

In an etching process with a very high rate constant the concentration of the reagent at the semiconductor surface may be so low that the diffusion of the reagent to the surface will be rate-determining. The rate of a diffusion-controlled etching process (r_d) is proportional to the concentration gradient at the surface. For one-dimensional diffusion with a linear concentration gradient it is given approximately by

$$r_d \approx \frac{D}{\delta} (c_b - c_s), \quad (11)$$

where D is the diffusion coefficient and δ the thickness of the 'diffusion layer' at the surface. The temperature dependence of the etch rate is now mainly determined by that of the diffusion coefficient, which is inversely proportional to the viscosity of the solution. For the aqueous solutions normally used we can derive an 'activation energy' of about 18 kJ/mol from the curve of $\log r_d$ against $1/T$.

If the rate of the surface reaction is of the same order of magnitude as the rate of diffusion, there are various methods for influencing the rate-determining step. In the case where the surface reaction is not of the first order ($n > 1$), it follows from eq. (9) that increasing the concentration c_b will result in a greater increase in the reaction rate, so that the diffusion becomes more important. A change in the same direction is obtained by increasing the temperature; see *fig. 9*. The surface reaction, which is rate-determining at low temperatures, is much more temperature-dependent than the diffusion, so that at higher temperatures the diffusion becomes the rate-determining quantity. A change in the opposite direction can be obtained by reducing the thickness δ of the diffusion layer; see

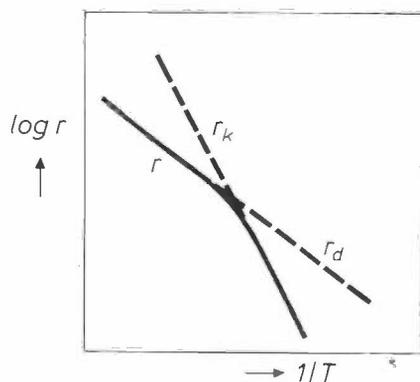


Fig. 9. Diagram showing the effect of the temperature T on the rate r and the kinetics of the etching process, where r_k is the etch rate when the surface reaction is kinetically controlled and r_d is the etch rate when the surface reaction is diffusion-controlled. Both $\log r_k$ and $\log r_d$ decrease linearly with $1/T$. At low temperatures, however, r_k is much smaller than r_d , so that the surface reaction determines the etch rate ($r = r_k$). At high temperatures r_k is much larger than r_d , and diffusion is then rate-determining ($r = r_d$).

eq. (11). This can be done by rotating the semiconductor or by stirring the solution. It is then possible to change a diffusion-controlled process into a kinetically controlled process.

Diffusion in electroless etching

In the example given in *fig. 6* the anodic partial current due to the oxidation of GaAs is kinetically controlled. The current depends only on the hole concentration at the surface. Near the mixed potential the reduction rate is determined by the diffusion of the oxidizing agent (Ce^{4+}) to the surface. This means that the rate of the total process is determined by diffusion via the reduction reaction. This situation is most frequently encountered in electroless etching.

We have been able to show that the etch rate can also be determined by diffusion via the oxidation reaction. This is illustrated in *fig. 10* for p-GaAs in a solution of 0.5 M $K_3Fe(CN)_6$ with a pH of 13. The total current-potential curve has three distinct plateaus. At low potentials the cathodic partial current due to the reduction of $Fe(CN)_6^{3-}$ is entirely determined by diffusion and does not depend on the potential. The reduction stops as soon as the potential approaches the redox potential of the $Fe(CN)_6^{3-}/Fe(CN)_6^{4-}$ couple. The anodic partial current due to the GaAs oxidation becomes important at the flat-band potential. At higher potentials, however, the anodic partial current also becomes constant. We have found that the etching process then depends on the transfer of OH^- ions to the semiconductor surface. The intermediate plateau in the total curve is the result of the difference between the potential-independent sections of the anodic and cathodic partial curves.

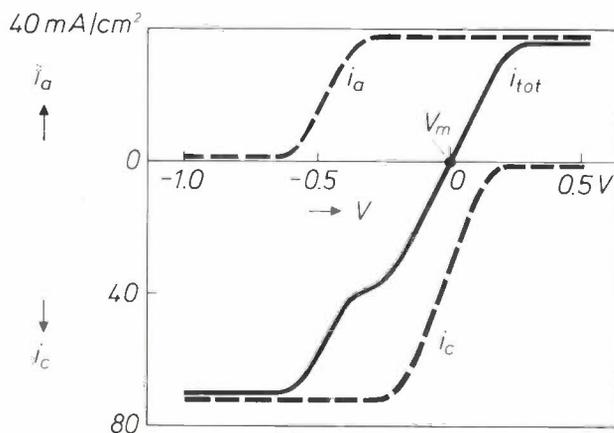


Fig. 10. Current-potential curves (as in *figs 6 and 7*) for p-GaAs in an 0.5 M $K_3Fe(CN)_6$ solution with a pH of 13. The anodic and cathodic partial current densities i_a and i_c are such that the curve for the total current density i_{tot} has three plateaus, corresponding to the constant value of i_c at low V , the constant value of i_a at high V and the difference between these two values. The etch rate, which is proportional to i_a at the mixed potential V_m , is not determined by a constant cathodic current in this case, but by a constant anodic current.

The constant cathodic current in fig. 10 is much higher than the constant anodic current. At the mixed potential it is not the cathodic current that is constant, but the anodic current, so that the etch rate is diffusion-controlled via the oxidation reaction. If the $K_3Fe(CN)_6$ concentration is decreased so as to make the constant cathodic current lower than the constant anodic current, the result, just as in fig. 6, is an etch rate that is diffusion-controlled via the reduction reaction.

Effect of the crystal planes

Semiconductors like GaAs, GaP and InP are normally used in single-crystal form. Their crystal structure corresponds to the zinc-blende structure of ZnS. The A atoms (Ga, In) form a face-centred cubic structure in which half the tetrahedral interstices are periodically filled with B atoms (As, P). Crystal planes such as the {100} and {111} planes can be distinguished from each other by a different density and arrangement of the atoms, which can affect the etch rate. Because of the polarity in the $\langle 111 \rangle$ directions, an A{111} plane, i.e. a {111} plane occupied only by A atoms, will behave differently from a B{111} plane^[12].

The arrangement of the atoms in the surface crystal plane can have a considerable effect on the flat-band potential V_{fb} . Table I gives the values of V_{fb} for the

Table I. Flat-band potential V_{fb} with respect to a saturated calomel electrode for three kinds of crystal planes of n- and p-GaAs in an NaOH solution with a pH of 14.

Crystal planes	V_{fb}	
	n-GaAs	p-GaAs
{100}	-1.80 V	-0.50 V
As{111}	-1.75	-0.45
Ga{111}	-1.55	-0.25

{100}, As{111} and Ga{111} planes of n-type and p-type GaAs. The difference in V_{fb} between corresponding crystal planes of n-type and p-type GaAs is approximately equal to E_g/e , where E_g is the band gap of GaAs (see p. 63). The {100} and As{111} planes have almost the same V_{fb} -value, but that of a Ga{111} plane is much higher. Partly because of this the anodic dissolution of a Ga{111} surface only starts at a potential more than 100 mV higher than the onset potential for dissolution of an As{111} surface.

In the literature various examples are given of the effect of the crystal planes on the rate of chemical etching, when this is not determined by mass transfer in the solution. It was found, for example, in the etching of InP^[13] that the rate varies in the sequence $P\{111\} \geq \{100\} \gg In\{111\}$. The explanation for this goes beyond the scope of this article^[12].

Etching at mask edges

If the etch rate is determined by diffusion for all crystal planes, isotropic etching would be expected at mask edges. Chemical etching does indeed give rounded profiles in this case. At a mask edge the etch depth is relatively large, because there is mass transfer both from the solution above the etched surface and from the solution beside the mask aperture. Fig. 11 shows a profile produced at a mask edge after a diffusion-controlled chemical etching process. It agrees well with the theoretical profile calculated from a two-dimensional model^[14].

If the etch rate of one of the crystal planes is determined by the rate of the surface reaction, facets are found at mask edges. The shape of the etch profile is then usually determined by the plane etched most slowly, which is usually an A{111} plane.

In electroless etching processes, profiles with facets may be observed even though the etch rate of all the

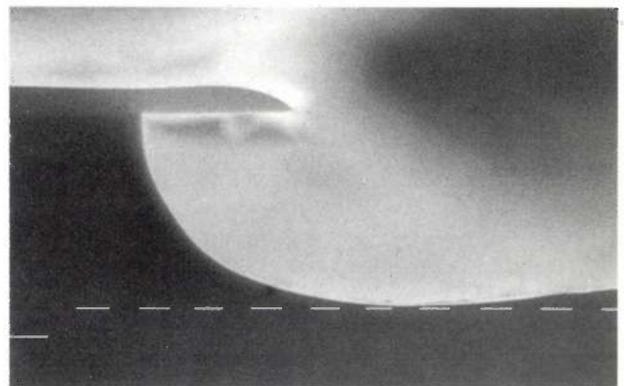


Fig. 11. Scanning electron-microscope photograph (scale divisions 1 μ m) of GaAs at a mask edge after a diffusion-controlled etching process with an H_2O_2/HCl solution. The etching results in a rounded profile, in good agreement with a profile calculated for this case^[14].

[12] See for example H. C. Gatos, Crystalline structure and surface reactivity, *Science* **137**, 311-322, 1962.

[13] R. Becker, Sperrfreie Kontakte an Indiumphosphid, *Solid State Electron.* **16**, 1241-1249, 1973.

[14] H. K. Kuiken, J. J. Kelly and P. H. L. Notten, Etching profiles at resist edges, I. Mathematical models for diffusion-controlled cases, *J. Electrochem. Soc.* **133**, 1217-1226, 1986; P. H. L. Notten, J. J. Kelly and H. K. Kuiken, Etching profiles at resist edges, II. Experimental confirmation of models using GaAs, *J. Electrochem. Soc.* **133**, 1226-1232, 1986.

individual planes is determined entirely by diffusion. This is related to the effect of cathodic protection, which we shall now briefly discuss.

Cathodic protection

When a metal is in contact with another metal during etching, the etch rates of the individual metals may be changed considerably by galvanic effects^[15]. The etch rate of the less-noble metal tends to increase, while the more-noble metal is etched more slowly. The more-noble metal is then said to be 'cathodically protected'.

This effect is also found in the electroless etching of different semiconductors or semiconductor regions in contact with each other^[16]. Fig. 12 gives a schematic representation of the current-potential curves for two p-doped III-V semiconductors, A_1B_1 and A_2B_2 , of equal area. Owing to the difference in 'nobility' the anodic partial curves of A_1B_1 and A_2B_2 are different. On the other hand the cathodic partial curves coincide, because the reduction rate is assumed to be determined by the diffusion rate of the oxidizing agent in the solution. Consequently the separate semiconductors are etched at the same rate, A_1B_1 at the mixed potential V_1 and A_2B_2 at the mixed potential V_2 . If

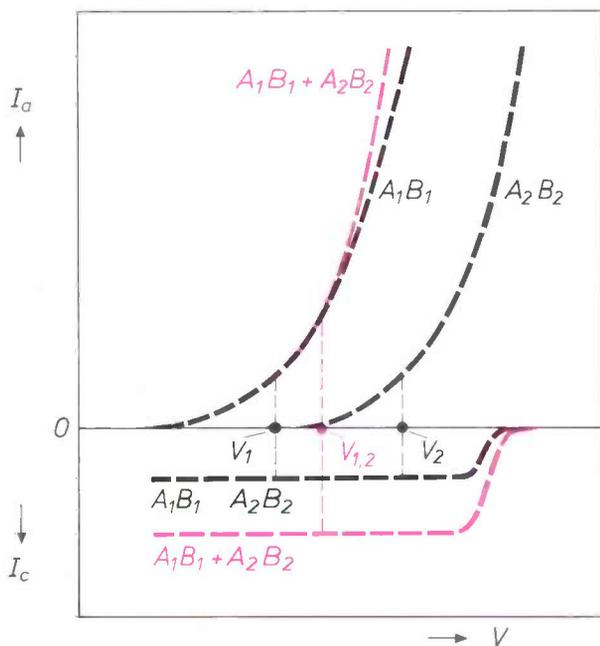


Fig. 12. Schematic current-potential curves illustrating the occurrence of cathodic protection in the etching of two semiconductors in electrical contact. Electroless etching of the individual semiconductors A_1B_1 and A_2B_2 at the mixed potentials V_1 and V_2 takes place at the same rate, determined by the constant value of the cathodic partial current. If there is electrical contact the partial currents of A_1B_1 and A_2B_2 have to be added together (red). Etching then takes place at an intermediate mixed potential, $V_{1,2}$, so that the 'less-noble' A_1B_1 is dissolved faster and the 'nobler' A_2B_2 more slowly.

A_1B_1 and A_2B_2 are in electrical contact, their partial current-potential curves have to be added together. This results in a new curve for the anodic process and in a cathodic current that is twice as large, because the areas of A_1B_1 and A_2B_2 are combined. This gives a new mixed potential ($V_{1,2}$), with a value higher than V_1 and lower than V_2 , which means that the etch rate of A_1B_1 is higher and that of A_2B_2 lower.

Cathodic protection can be useful for selectively etching semiconductors in multilayer structures: while there is little or no difference in the etching behaviour of the separate materials, there can be a considerable difference in etch rate because of the electrical contact between them.

Practical examples

The applications of wet-chemical etching of III-V semiconductors are much too numerous for a detailed description of each. We shall therefore confine ourselves to a few interesting applications that we have investigated.

Material quality control

Although the semiconductors used for making electronic devices are true single crystals, they always have some crystallographic defects. These defects, which may decide the quality and life of a component, can be revealed by a treatment with a defect-selective etchant. Defects usually correspond to crystallographically perturbed areas such as dislocations and stacking faults. Since these areas will behave less 'nobly' than their environment, they can be selectively etched in the dark. The higher etch rate at defects leads to the formation of etch pits. The sensitivity of etchants based on this principle is in general low: to make all the defects visible it is necessary to etch away more than 10 μm . Obviously this cannot be done for thin films.

Crystallographic defects can also be revealed by photo-etching, since it is known that effective recombination of free electrons and holes often takes place at such defects. Since the photo-etch rate depends on the surface concentration of both types of charge carrier, the increased recombination at crystal defects produces a local decrease in the etch rate. Photo-etching therefore makes the sites with defects stand out from their environment ('hillocks'). The high sensitivity that can be achieved in this way is demonstrated in fig. 13, which shows a photomicrograph of n-GaAs after photo-etching with an $\text{H}_2\text{O}_2/\text{H}_2\text{SO}_4$ solution. Many defects are visible after removal of 0.4 μm of the surface. Striations due to areas of different doping concentration also appear, as parallel lines.

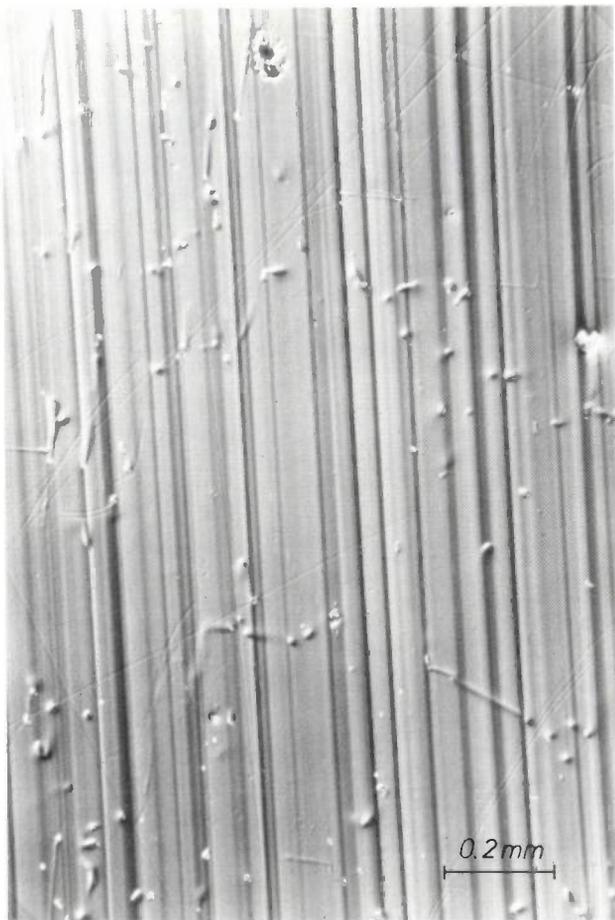


Fig. 13. Photomicrograph (interference contrast) of n-GaAs after photo-etching with an $\text{H}_2\text{O}_2/\text{H}_2\text{SO}_4$ solution for revealing crystal defects^[8].

The CrO_3/HF solutions mentioned earlier are also suitable etchants for the quality control of III-V semiconductors. In these etchants the crystallographic defects are revealed by the formation in the dark of a thin passivating film, which is slightly thicker at the defects than on a perfect surface^[7]. The defects are thus preferentially passivated, so that they are etched away more slowly. In n-GaAs and n-InP the sensitivity can be increased further by illumination; this effect can again be attributed to increased recombination of electrons and holes at the defects.

Profile etching

The etching of special shapes and profiles with the aid of a resist is a very important process in III-V semiconductor technology. To take a particular example, structures with symmetrically rounded edges or grooves with a V-shaped cross-section are frequently used for semiconductor lasers. Good use can often be made here of etching kinetics. As we have shown, etching processes with a rate-determining surface reaction give structures with facets, whereas the

diffusion-controlled chemical etching processes are isotropic. We have also shown that the etching kinetics can be changed, for example by raising the temperature (fig. 9). This makes it possible to influence the etching profile fairly easily. Fig. 14 shows an example in which the profiles were made in GaAs by chemical etching at different temperatures. At 0°C the surface reaction is still rate-determining and a V-groove is produced. At 20°C diffusion becomes important,

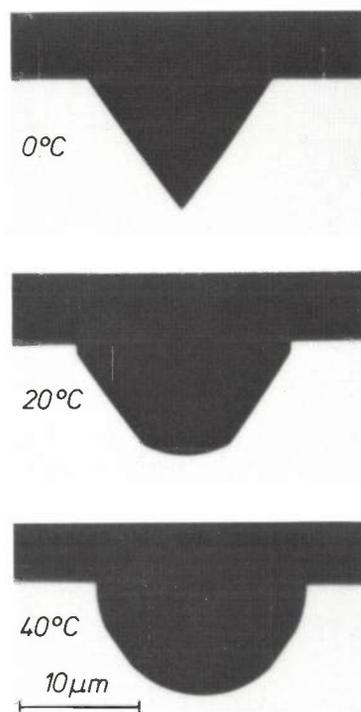


Fig. 14. Photomicrographs of profiles in GaAs after chemical etching with a Br_2/KBr solution at different temperatures. A change in temperature has a pronounced effect on the etching kinetics and gives a completely different etching profile: from a V-groove at 0°C to an almost completely rounded profile at 40°C .

with the result that the profile is slightly rounded. At 40°C diffusion is rate-determining and the result is an almost completely rounded profile.

Another factor that has an important bearing on the shape of the etching profile is the orientation of the mask in relation to the crystal orientation. If for example a narrow track in the $[\bar{1}10]$ direction is etched into a (001) surface, the result is a channel with a V-shaped cross-section, formed by two $\{111\}$ planes

^[15] J. J. Kelly and C. H. de Minjer, An electrochemical study of undercutting during etching of duplex metal films, *J. Electrochem. Soc.* **122**, 931-936, 1975;

J. J. Kelly and G. J. Koel, Galvanic effects in the wet-chemical etching of metal films, *Philips Tech. Rev.* **38**, 149-157, 1978/79.

^[16] H. Löwe and I. Barry, Zur Kinetik der Germaniumauflösung in alkalischen Hexacyanoferrat-III-Lösungen, *Z. Phys. Chem., Leipzig* **249**, 73-80, 1972.

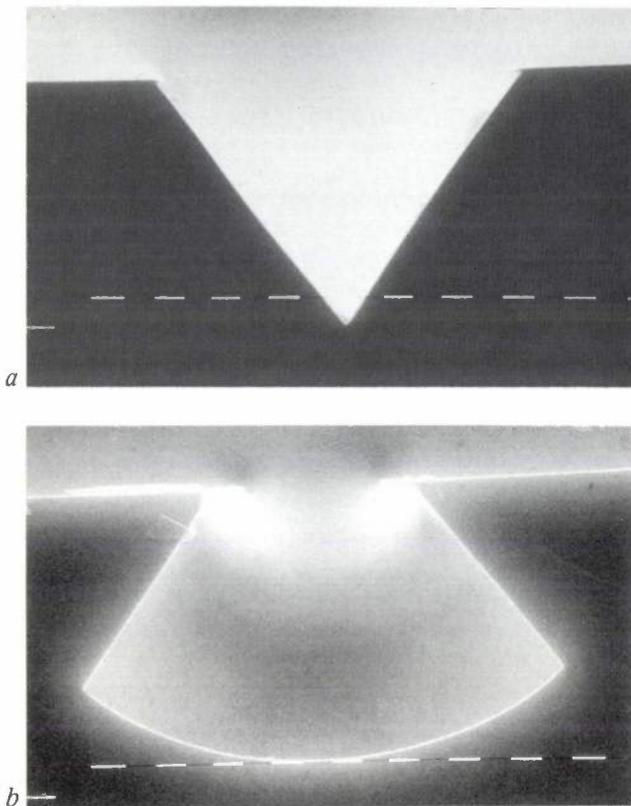


Fig. 15. Scanning electron-microscope photographs (scale divisions 1 μm) of GaAs after chemical etching with a Br_2/HBr solution for two different mask orientations. When a narrow track is etched in the $[110]$ direction a V-groove is formed (a). When this is done in the $[110]$ direction, the result is a 'dovetail' (b).

with A atoms; see *fig. 15a*. The characteristic etching angle between the side planes and the surface is 55° , corresponding to the angle between the (001) plane and the $\{111\}$ planes. If on the other hand a narrow track is etched in the $[110]$ direction (i.e. perpendicular to the $[\bar{1}10]$ direction), then the etching spreads outwards with a characteristic angle of 125° , corresponding to the supplement of the angle between the (001) plane and the $\{111\}$ planes. This results in a channel with a 'dovetail' cross-section (*fig. 15b*).

In electroless etching the etch rates for the various crystal planes are clearly often diffusion-controlled, and yet profiles with facets are obtained. In these cases the etch rate is determined by the cathodic partial current due to the reduction reaction (figs 6 and 7). During etching at a mask edge various crystal planes are exposed to the etchant. As we have shown, these have different V_{fb} -values (Table I) and the anodic-current-potential curve depends on the crystal plane. This means that galvanic effects can occur between the planes of the same crystal. In an analogous way to that shown in *fig. 12*, the most-noble plane is cathodically protected by the rest of the surface, and this results in the formation of a facet.

This effect does not occur when, as in *fig. 10*, the etch rate is determined by a diffusion-controlled anodic current. This situation is shown schematically in *fig. 16*. In this case the diffusion-controlled cathodic current (not shown) is very high and there are two planes of quite different nobility. It can be seen from the anodic current at the mixed potential that the electrical contact between the planes has no effect on their etch rate. There is no cathodic protection, so that the planes dissolve at the same diffusion-controlled rate and a rounded profile is obtained.

In the photo-etching of III-V semiconductors using light of very short wavelength (e.g. 350 nm) an exceptionally high anisotropy can be achieved^[9]. The strong absorption of the light induces a relatively high concentration of holes in the exposed part of the semiconductor surface, resulting in a highly direction-dependent etching process. Since the etched cavity acts as a waveguide for the incident light, and dissolution only occurs at the bottom of the cavity, structures with very deep holes or grooves can be produced. Structures of this type offer interesting prospects for practical application.

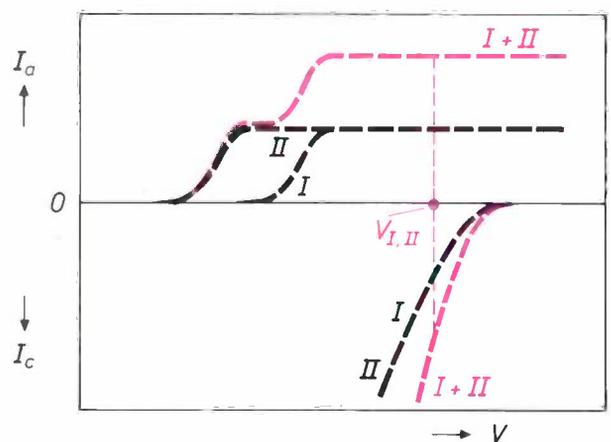


Fig. 16. Schematic current-potential curves for the etching of two crystal planes *I* and *II*, of dissimilar 'nobility', with no cathodic protection. Both planes give the same diffusion-controlled constant anodic current, which is much lower than the diffusion-controlled cathodic current (not shown here). At the mixed potential $V_{I,II}$ the total anodic current (red) is equal to the sum of the constant anodic currents of *I* and *II*, so that the electrical contact between *I* and *II* has no effect on their etch rates.

Selective etching

In III-V semiconductor technology there is a need for selective etchants that are highly sensitive to the composition of the material. High selectivity is particularly important for etching thin-film structures, as in semiconductor lasers, in which the active layer is no thicker than $0.1 \mu\text{m}$. To illustrate what can be achieved in this field we shall discuss a few typical examples.

With chemically dissimilar materials the required selectivity can often be obtained by using the appropriate etchant. For instance, InP can be dissolved selectively with respect to $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ by chemical etching in a concentrated HCl solution, in which the $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ is hardly attacked at all. Fig. 17 shows a multilayer structure with these semiconductors, in which the InP has been partly removed by selective etching. The opposite selectivity can be obtained in electroless etching, e.g. with a $\text{Ce}^{4+}/\text{H}_2\text{SO}_4$ solution. The $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ readily dissolves in this solution, while the InP is passivated, probably because of the formation of an oxide film on the surface. An example of a multilayer structure etched in this way is shown in fig. 18.

Both forms of selectivity are also possible with GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$. In etching multilayer structures with these materials the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is pas-

sivated to such an extent in some solutions, by the formation of an oxide film, that virtually only the GaAs is attacked; see fig. 19. A suitable method for selectively dissolving $\text{Al}_x\text{Ga}_{1-x}\text{As}$ with respect to GaAs is based on the difference in the nobility of these compounds in electroless etching. Fig. 20 shows a multilayer structure treated by this method. Etching through a mask first produces a narrow hole in the upper GaAs layer, and then a much wider hole in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layer beneath it. The bottom GaAs layer is not attacked, however, because of its greater nobility and the cathodic protection.

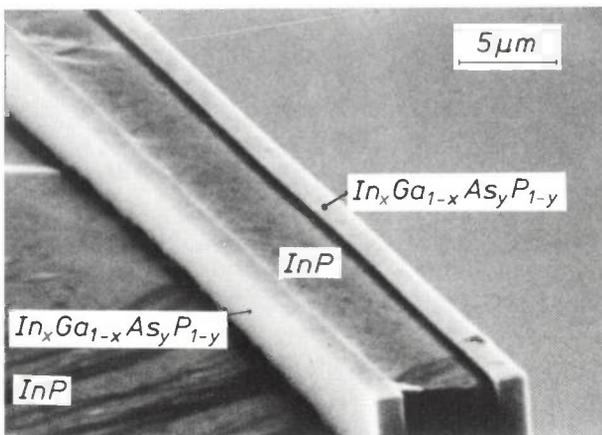


Fig. 17. Scanning electron-microscope photograph of a multilayer structure of InP and $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$, after chemical etching with a concentrated HCl solution. Only the InP layers have been partly etched away.

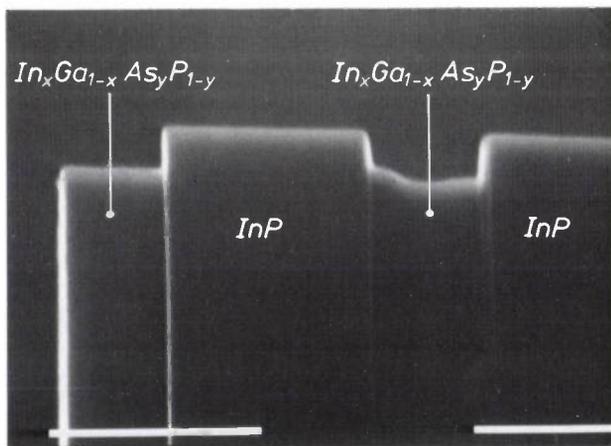


Fig. 18. Scanning electron-microscope photograph (scale divisions 1 μm) of a multilayer structure of InP and $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$, in which electroless etching with a $\text{Ce}^{4+}/\text{H}_2\text{SO}_4$ solution has only affected the $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$.

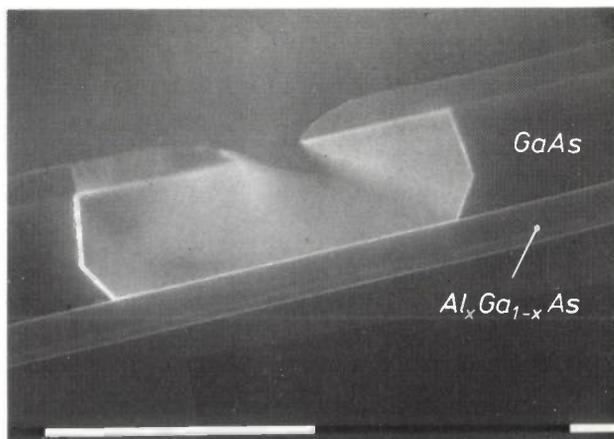


Fig. 19. Scanning electron-microscope photograph (scale divisions 1 μm) of a multilayer structure of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$, in which only GaAs has been partly removed by chemical etching with an H_2O_2 solution.

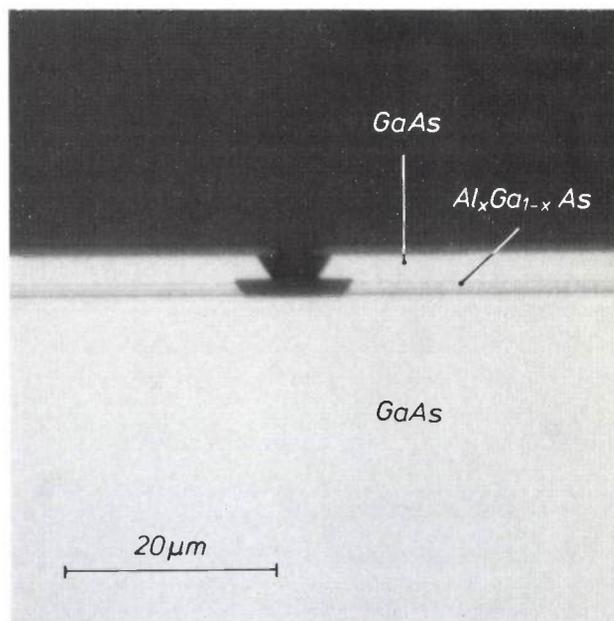


Fig. 20. Photomicrograph of a multilayer structure of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$. Electroless etching through a mask has produced a relatively small hole in the top GaAs layer and a much wider hole in the underlying $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layer. The bottom GaAs layer has not been attacked, however.

Ned. Philips Bedrijven B.V.
 PHILIPS RESEARCH LABS.
 LIBRARY WY - 1
 P.O. Box 80.000
 5600 JA EINDHOVEN
 THE NETHERLANDS

The selective etching of semiconductor materials that differ only in the type of doping is in general not possible by electroless or chemical etching. Although some selectivity may be obtained with electroless photo-etching, the most reliable method is based on anodic etching. The principle can be simply illustrated with reference to fig. 3. The p-GaAs is selectively dissolved in the dark at a potential near its flat-band potential (about 0.3 V), whereas on exposure to light and with a lower potential only the n-GaAs is dissolved. In this way regions on either side of a pn-junction in GaAs can be etched selectively. A p-doped InP layer can also be completely removed from an n-InP layer beneath it by anodic etching.

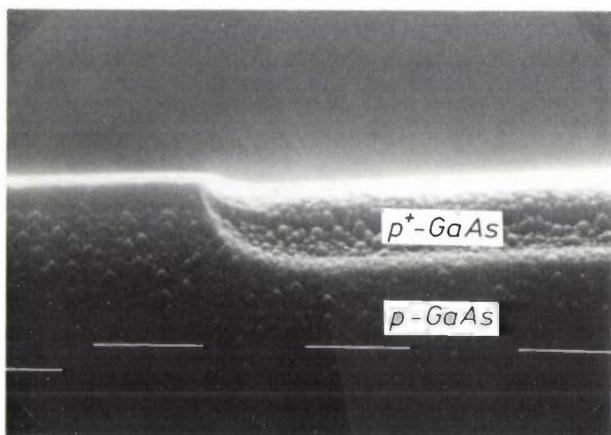


Fig. 21. Scanning electron-microscope photograph (scale divisions 1 μm) of a GaAs layer in which the heavily doped p^+ region has been etched away with a $\text{K}_3\text{Fe}(\text{CN})_6$ solution; the less heavily doped p region has hardly been attacked.

By using a difference in potential distribution, with cathodic protection, we have succeeded in selectively etching semiconductor materials that differ only in doping concentration. Such combinations of materials are found in laser structures with a top layer of p-GaAs that is heavily doped (p^+) locally to give good electrical contact. The contact region can be revealed by selective etching with a solution of 0.1 M $\text{K}_3\text{Fe}(\text{CN})_6$ at a pH of 14; see fig. 21. The only difference between the p and p^+ regions shown here is the doping concentration: 10^{18} and 10^{20} cm^{-3} . Because of the near-metallic nature of the p^+ region, the potential distribution at the interface with the solution is quite different from that for a conventional p-type semiconductor. The p^+ region is less noble and is more rapidly dissolved when in contact with a large p-type area.

Summary. III-V semiconductors like GaAs can be wet-chemically etched by three mechanisms: electrochemically with an external voltage source, electrochemically using an oxidizing agent (electroless), and chemically with a reactive compound. In some cases the etching process only proceeds when the semiconductor is exposed to light. The etch rate depends on the relative reaction rate at the semiconductor surface and the mass transfer in the solution. Other important factors are the effect of the crystal planes, the orientation of a mask with respect to these planes, and the electrical contact with other materials. Wet-chemical etching of III-V semiconductors can be used on a large scale for various applications, including the detection of crystallographic defects, the fabrication of special profiles and the selective dissolution of closely related materials in multilayer structures.

1938

THEN AND NOW

1988

Lamps for phototherapy

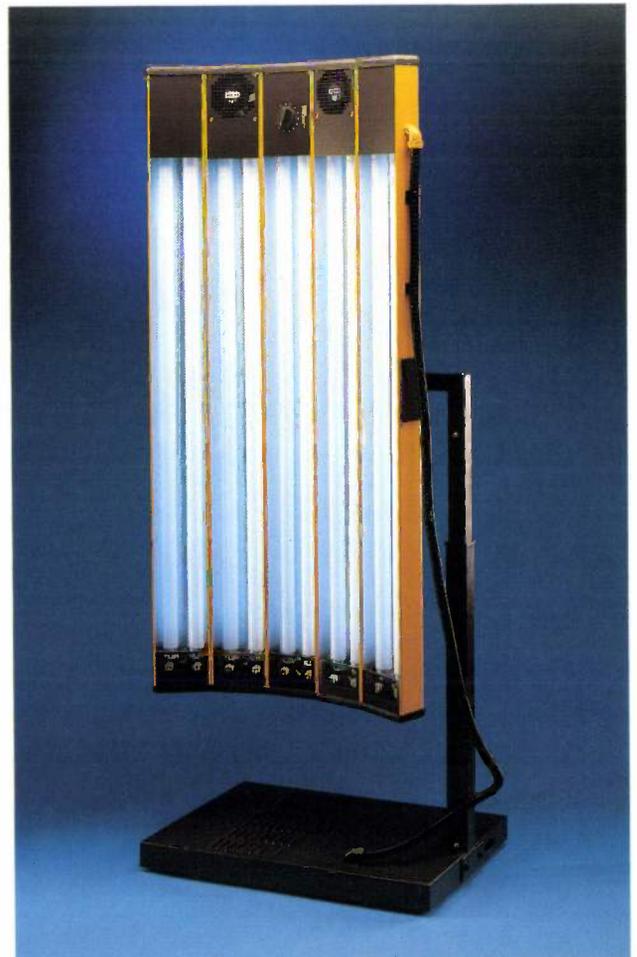


The treatment of certain complaints by radiation with artificial light (phototherapy) first became known in about 1900. The method grew in importance as new kinds of lamps were invented and as promising results were obtained with solar radiation. Artificial light freed the therapist from the uncertainties of the weather, and the radiation could be kept to a narrow wavelength range to give a better therapeutic action with fewer harmful side effects.

It is about fifty years since Philips developed the 'Biosol' lamp for ultraviolet radiation^[*] (black-and-white photograph). This was a tubular discharge lamp with a length of about 20 cm, a gas pressure of 1 bar and a power of 250 or 475 W. The lamp was intended for the treatment of diseases such as rickets, with radiation at a wavelength of about 250 nm, and lupus (tuberculosis of the skin), at a wavelength of 320 to 350 nm. Filters gave a wavelength distribution as close to the ideal as possible.

With the passage of time, progress in lighting technology and our improved understanding of phototherapy

have enabled us to match the wavelength more closely to the ailment. One important step was the application of fluorescent lamps, in which a phosphor converts the ultraviolet radiation originating in the discharge into radiation at a longer wavelength. For the treatment of the skin complaint psoriasis, for example, Philips developed the TL/12 fluorescent lamp, which has a relatively wide emission band centred on 305 nm. After it had been found that the therapeutic effect increased as the wavelength of the radiation approached 315 nm, a new type of lamp was designed — the TL/01. Experiments with this lamp are under way in several Dutch hospitals and in other countries. The colour photograph shows an experimental array of ten of these lamps. They are 1.20 m long, each has a power of 40 W and they have a very narrow emission band with a maximum at 311 nm. The first results from these experiments look very promising.



[*] From Philips Technical Review, January 1938.

Bistability in quantum-well lasers

A. I. Kucharska, P. Blood and E. D. Fletcher

Since the early sixties there has been a growing interest in using optical nonlinearities of semiconductor materials for optical logic. Amongst a variety of devices particular attention has been given to the operation of inhomogeneously pumped injection lasers showing two stable light-output levels at the same injection current. Recent advances in the techniques for semiconductor-layer growth have made possible the fabrication of lasers with very thin active layers known as quantum wells. The article below presents some preliminary results of an investigation of the optical bistability in inhomogeneously pumped quantum-well lasers.

Introduction

An inhomogeneously pumped semiconductor injection laser may show optical bistability in the form of two stable light-output levels for the same value of injection current^[1]. This nonlinear optical effect relies on saturation of the light absorption in a passive region in the device. It offers the possibility of making optical switching devices to be used for modulating light sources and processing information which is being transmitted or manipulated by means of light beams.

In recent years there has been particular interest in the multiple-quantum-well (MQW) laser, where the light is generated in a set of GaAs layers which are thinner than 20 nm and which are sandwiched between layers of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ with a larger band gap^[2]. In the active layers the motion of electrons and holes perpendicular to the interfaces is quantized. In addition to fascinating physical phenomena, MQW lasers show some practical advantages such as a relatively low threshold current and a short emission wavelength^[3]. It is also expected that the nonlinear optical effect due to inhomogeneous pumping is stronger in MQW lasers than in lasers having a thick active layer

of 'bulk' GaAs. The ability to make use of this enhanced nonlinearity could lead to the development of highly efficient electro-optic logic elements.

The fabrication of MQW lasers requires advanced growth techniques such as molecular beam epitaxy (MBE), which has become highly developed at Philips Research Laboratories in Redhill, England^[4], or metal-organic vapour-phase epitaxy. MBE is a refined form of vacuum evaporation in which molecular (or atomic) beams from effusion cells impinge upon a heated substrate under ultra-high vacuum conditions. The layer-growth rate for a material like GaAs is

^[1] See for example M. J. Adams, A tentative assessment of semiconductor laser optical bistability, *Int. J. Electron.* **60**, 123-142, 1986.

^[2] Quantum-well structures have recently been described in this journal, in a special issue on layered semiconductor structures, *Philips Tech. Rev.* **43**, 109-165, 1987.

A description of the theoretical aspects of the quantum well will be given by R. Eppenga and F. M. H. Schuurmans, *Philips Tech. Rev.* **44**, No. 5, Autumn 1988.

^[3] K. Woodbridge, P. Blood, E. D. Fletcher and P. J. Hulyer, Short wavelength (visible) GaAs quantum well lasers grown by molecular beam epitaxy, *Appl. Phys. Lett.* **45**, 16-18, 1984; P. Blood, E. D. Fletcher and K. Woodbridge, Dependence of threshold current on the number of wells in AlGaAs-GaAs quantum well lasers, *Appl. Phys. Lett.* **47**, 193-195, 1985.

^[4] B. A. Joyce and C. T. Foxon, Molecular beam epitaxy of multilayer structures with GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$, *Philips Tech. Rev.* **43**, 143-153, 1987.

typically about 1 μm/h, i.e. one monolayer per second, so with beam shutters in front of each cell operating within 0.3 s it is possible to grow very thin layers with interfaces that are ‘abrupt’ within one monolayer.

In the investigation described in this article, we have performed some preliminary experiments on inhomogeneously pumped quantum-well laser structures in the GaAs/Al_xGa_{1-x}As material system grown by MBE, and the measured light-current characteristics do indeed show hysteresis. It has been possible to demonstrate rapid switching between low and high light-output levels. The mechanism of the optical bistability in MQW lasers was assessed by studying the laser-emission spectrum and the optical absorption spectrum of the passive region.

In this article we first give some details of the MBE-grown laser structures. Next we describe their light-

K. Woodbridge using a laboratory-built MBE system with computer-controlled effusion-cell temperatures and shutter operations [3]. The structure has a 200 nm thick waveguiding region of Al_{0.35}Ga_{0.65}As barriers and two 2.5 nm thick quantum wells of GaAs. On both sides cladding layers of Al_{0.60}Ga_{0.40}As were grown. Contact layers of heavily doped n-GaAs and p-GaAs were used to obtain good electrical contacts to the metallizations. A stripe laser, 50 μm wide and 300 μm long, was made by P. J. Hulyer using oxide insulation. The laser operated at a wavelength of 777 nm. Inhomogeneous pumping was achieved by dividing the p contact of the device into two regions with a groove etched through the top metallization and top GaAs contact layer in a direction perpendicular to the oxide stripe. It was thus possible to have a pumped region and an unpumped region within the same Fabry-Pérot cavity.

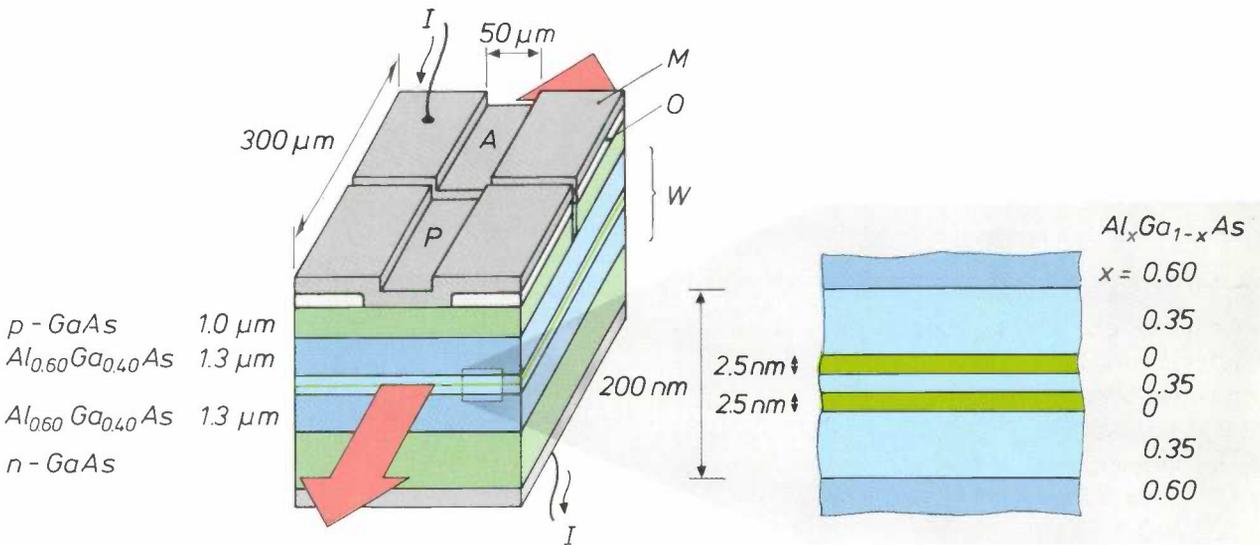


Fig. 1. Schematic diagram of the structure of an inhomogeneously pumped MQW laser device. The device has a length of 300 μm and the current *I* is restricted to a 50 μm wide stripe by insulation with an oxide layer *O*. The laser light *L* is generated in two 2.5 nm thick quantum wells of GaAs sandwiched between layers of Al_{0.35}Ga_{0.65}As. The waveguide *W* is formed by two 1.3 μm thick cladding layers of Al_{0.60}Ga_{0.40}As. Contact layers of heavily doped n-GaAs and p-GaAs are used for the electrical contact with the metal layers *M*. A groove perpendicular to the stripe has been etched through the top metal layer and top contact layer, resulting in an active region *A* and a passive region *P* for laser operation.

current characteristics and switching behaviour. Finally, the physical processes responsible for the observed bistability are discussed.

Device structures

The structure of an inhomogeneously pumped MQW laser device is shown schematically in *fig. 1*. The epitaxial semiconductor structure was grown by

Light-current characteristics and switching

The light-current characteristics were measured in real time by driving one contact segment of the devices with a triangular current pulse, with an overall duration of 800 ns and a repetition rate of 1 kHz, and by detecting the generated light with a fast avalanche photodiode. For the inhomogeneously pumped devices curves with a hysteresis were obtained as shown in *fig. 2* for a typical MQW device. As current is first

injected the light output increases slowly. At higher current the light output increases steeply when the laser operation switches on. If in this state the current is decreased slightly, the laser remains 'on'. When the current is still further decreased the light output suddenly drops to a low level. It is then necessary to drive the device to a higher current in order to turn it 'on' again. The difference between the threshold current for laser action and the turn-off current was not affected by the maximum current, I_{\max} , to which the device was driven above threshold. This is also shown in fig. 2 for two different maximum current values, I_{\max}^a and I_{\max}^b .

The hysteresis observed in the light-current characteristics of these structures suggests that it should be possible to cause them to switch between two stable optical power levels. This was demonstrated by driv-

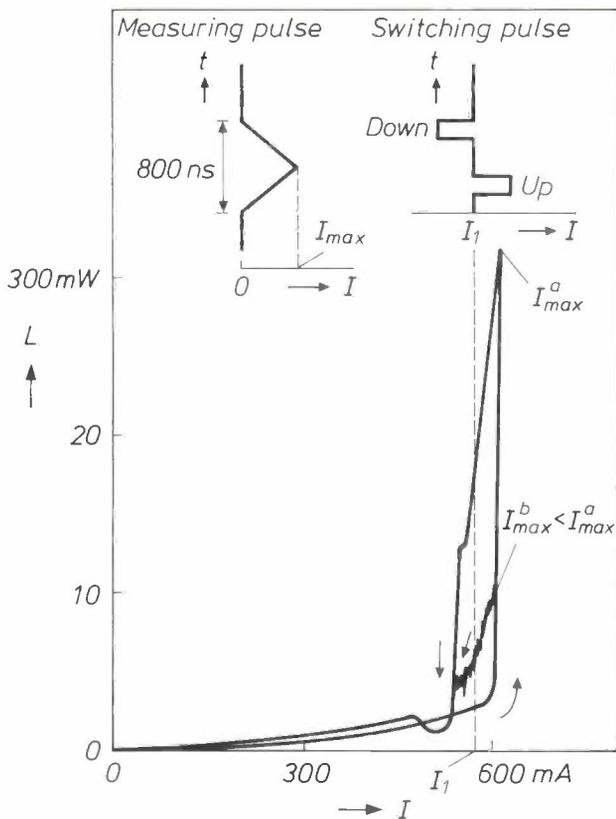


Fig. 2. Light-current characteristic of an inhomogeneously pumped MQW laser device showing hysteresis. Also shown are the forms of the current pulses (current I against time t) used for this measurement and a switching experiment. The device had a total length of $324 \mu\text{m}$ and the ratio of the pumped and unpumped stripe lengths was approximately 4:1. The current was applied as triangular pulses of 800 ns duration with a maximum value I_{\max} above the threshold current for laser action. The turn-off value of the current in the hysteresis loop was not affected by this maximum value (I_{\max}^a or I_{\max}^b). Because of the hysteresis, two light-output levels exist at the current I_1 . As indicated, switching-up and switching-down between these two levels is made possible by applying a short pulse giving a slight increase or decrease of the current with respect to I_1 (see also fig. 3).

ing the devices with a relatively long pulse ($\approx 500 \text{ ns}$) to a current I_1 (fig. 2) within the hysteresis loop, then superimposing shorter ($\approx 20 \text{ ns}$) switch-up and switch-down pulses having amplitudes which drove the devices above threshold current and below the turn-off current respectively. Fig. 3 shows typical oscilloscope traces of the monitored device current and the light output as functions of time. Switching is clearly demonstrated with the light output remaining at a high

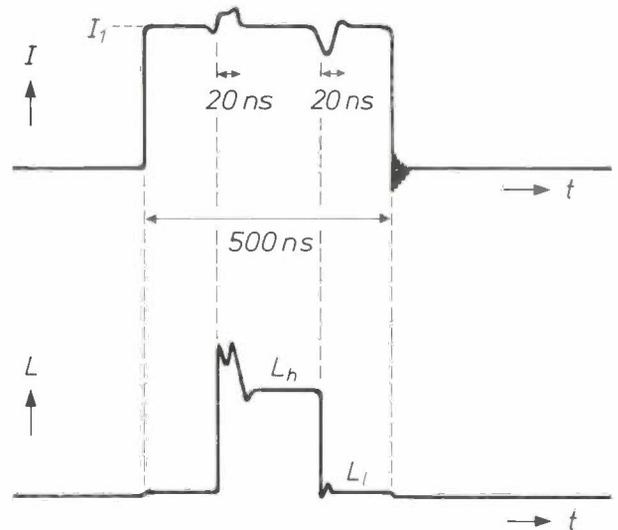


Fig. 3. Current I and light output L against time t showing the switching behaviour of an inhomogeneously pumped MQW laser device. At the current I_1 the light output is either at a high level (L_h) or at a low level (L_l) depending on whether the preceding short (20 ns) switching pulse was upwards or downwards with respect to I_1 . The expert help of A. Gowthorpe with the electronics for the pulsed experiments is gratefully acknowledged.

level after the switch-up pulse was removed and the current returned to I_1 . Likewise, the output remained at a low level for the same base current I_1 after the switch-down pulse was applied. Thus at the current I_1 , the light output is at one of two stable levels according to the sense of the preceding short switching pulse. The same type of hysteresis and switching was observed for a variety of inhomogeneously pumped MQW devices made from the same slice. The rise time of the light output in response to a switch-up pulse was less than 2 ns , our present experimental limit, and the optical power at the threshold current was a few milliwatts (fig. 2). The hysteresis described here with the associated switching behaviour was not observed for homogeneously pumped structures made from the same materials.

This kind of switching is due to the passive region in the laser cavity, and we have observed similar behaviour in conventional double-heterostructure

lasers which do not embody quantum wells. We argue below, however, that the mechanism in quantum wells is different from that in bulk materials, and one of the objectives of our research is to determine whether quantum wells offer advantages in these devices.

Mechanism of the bistability

We believe that the observed hysteresis in the light-current characteristics of the inhomogeneously pumped devices is due to reduction of the optical absorption within the unpumped region as the optical power is increased, as in other devices of this type^{[1][5]}. The importance of the optical absorption can be illustrated by emission spectra measured from each end of an inhomogeneously pumped device at different currents below threshold; see fig. 4. Since the light absorption within the cavity is stronger at the short-wavelength side, it has the effect of moving the spontaneous-emission peak to longer wavelengths. In our device, the spontaneous emission is generated only within the pumped region. Consequently, for a single pass through the cavity, this absorption has a more pronounced effect on the spectrum from the unpumped end as no light is generated near this facet. This explains why the spontaneous emission peak from the unpumped end occurs at longer wavelengths than that from the pumped end (fig. 4).

As the current is increased, the gain peak moves to higher energy, thus the emission peak moves to shorter wavelengths. Eventually there is sufficient gain in the pumped region to overcome the absorption losses in the unpumped region over some part of the spectrum. Thus an additional peak appears on the long-wavelength side of the spectrum emitted from the pumped end. This peak coincides with the emission peak from the unpumped end at the threshold current, and represents light which is amplified over several round-trips of the cavity, leading to laser action at threshold.

From the fast rise time of the light output in response to a switch-up pulse we infer that the optical bistability is caused by electronic processes rather than thermal processes. In bulk GaAs, experiment has shown that the effect of an increasing optical power on the absorption depends on the photon energy of the light with respect to the band gap of the semiconductor material^[6]. For light having a photon energy lower than the original band gap, the absorption increases due to a reduction of the band gap as a result of Coulomb interactions between the increasing number of charge carriers. For light having a photon energy higher than the original band gap, the absorption decreases with increasing optical power as band

filling reduces the probability of valence-to-conduction band transitions because of the Pauli exclusion principle. This effect most probably causes the optical bistability in conventional double-heterostructures with an active layer of bulk GaAs.

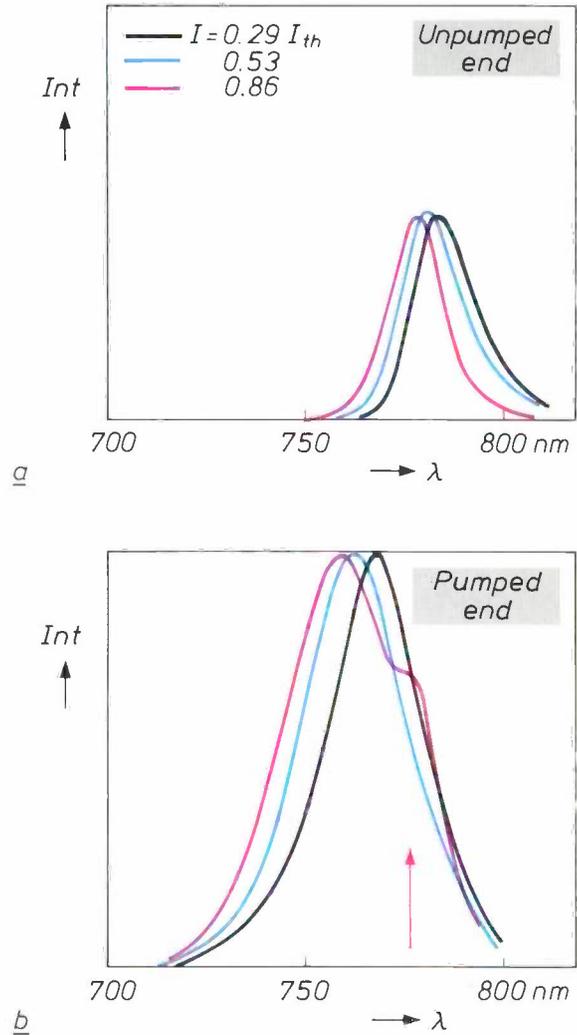


Fig. 4. Normalized emission spectra from the unpumped end and the pumped end of an inhomogeneously pumped MQW laser device. The relative intensity Int (in arbitrary units) is plotted against the wavelength λ of the emission for three values of the injection current I below the threshold current (I_{th}) for laser action. The emission peaks for the unpumped end have a lower intensity and are situated at longer wavelengths than those for the pumped end. At increasing injection the maxima shift to shorter wavelength. At $I = 0.86 I_{th}$ an additional peak appears in the spectrum from the pumped end at about the same wavelength as the maximum for the unpumped end, coinciding with the laser-emission wavelength (arrow) at $I \geq I_{th}$.

- [5] S. Tarucha and H. Okamoto, Voltage-controlled optical bistability associated with two-dimensional exciton in GaAs-AlGaAs multiple quantum well lasers, *Appl. Phys. Lett.* **49**, 543-545, 1986.
- [6] Y. H. Lee, A. Chavez-Pirson, S. W. Koch, H. M. Gibbs, S. H. Park, J. Morhange, A. Jeffery, N. Peyghambarian, L. Banyai, A. C. Gossard and W. Wiegmann, Room-temperature optical nonlinearities in GaAs, *Phys. Rev. Lett.* **57**, 2446-2449, 1986.

In MQW devices, however, the absorption at the band edge may be strongly affected by the presence of bound electron-hole pairs known as excitons^[7]. At low carrier densities excitons appear much more strongly in MQW systems than in bulk material due to the effects of spatial confinement which increase the probability that the motion of an electron and a hole will be correlated to form an exciton, and increase the binding energy of such an electron-hole pair. Thus large resolved exciton peaks are observed in the absorption spectra of MQW structures even at room temperature. The active region of the device used in this work was sufficiently lightly doped (10^{16} cm^{-3}) for excitons to be present in unpumped material. At high carrier densities, however, the excitons are screened out by many-body interactions and the absorption saturates, giving rise to a large optical nonlinearity at these energies.

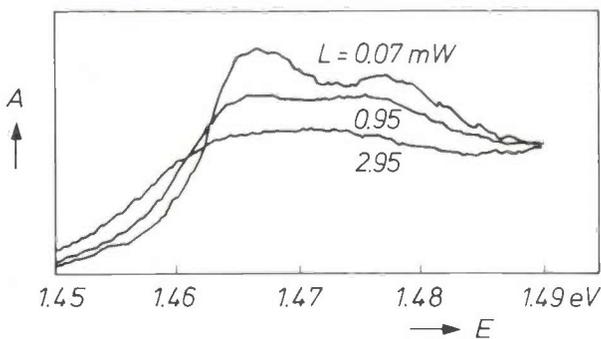


Fig. 5. Room-temperature absorption spectra of an MQW structure having a hundred 8.5-nm quantum wells, measured with the light propagating perpendicular to the wells. The sample was grown by MBE in the same system as the bistable devices. The absorption A , measured at three different levels of the incident optical power L , is plotted in arbitrary units against the photon energy E . Whereas the absorption increases with increasing power at the low-energy side ($\approx 1.45 \text{ eV}$), it decreases at the energy where excitonic absorption occurs. The data for this figure were provided by A. Miller, of the Royal Signals and Radar Establishment, Great Malvern, England.

Fig. 5 shows the absorption spectrum of an MQW structure having 8.5 nm thick quantum wells, grown by MBE in the same system as the present bistable device structures, measured for three optical-power levels with the direction of light propagation perpendicular to the wells. It can be seen that the exciton peaks present at low power levels are screened out with increasing optical intensity, while the simultaneous effect of band-gap reduction is to bring the con-

tinuum absorption edge to lower energies. Thus the net result is that the absorption increases at low energies, but decreases in the vicinity of the exciton peaks. Above the original band edge the absorption may also decrease due to band filling. In our case, the inhomogeneously pumped MQW devices with 2.5-nm wells operate below the original band edge in the region of the excitonic peaks in the absorption spectrum. Thus the decrease in absorption was provided by the excitons and not by the continuum. The subsidiary peak in the emission spectrum of fig. 4, at which the device switches, thus arises from the decreasing losses at the exciton resonance.

An additional effect which could reduce the absorption below the band edge in inhomogeneously pumped MQW structures is the quantum-confined Stark shift of the band edge under the influence of an electric field. This shift arises from the change in the shape of the potential distribution in a quantum well when an electric field is applied across it. The potential changes in such a way that the energy difference between the lowest quantum states of the electrons and holes is reduced^[8]. In the unpumped segment of the bistable MQW laser structure there is a built-in field due to the presence of the p-n junction. This field is reduced at high injection current by carriers generated upon absorption of light from the pumped segment. This reduces the Stark shift, so that the absorption edge moves to a higher energy at a higher injection current. We have observed a shift of the absorption edge of this structure of approximately 3-4 meV for an applied bias of 9 V ($6.8 \times 10^5 \text{ V/cm}$) whereas the built-in field is only about $1.8 \times 10^5 \text{ V/cm}$. Even if this field is reduced to zero under optical excitation, the resulting shift is less than about 1 meV, which is too small to be responsible for the bistability we observe.

In conclusion, we believe that the bistability of the inhomogeneously pumped MQW devices is due to excitonic saturation. Experiments are planned to investigate this further and to compare the performance of MQW and bulk materials in these switching devices.

Summary. Hysteresis has been observed in the light-current characteristics of inhomogeneously pumped GaAs-Al_xGa_{1-x}As laser structures which were grown by molecular beam epitaxy and whose active regions consisted of 2.5 nm thick multiple quantum wells (MQW). Fast switching ($< 2 \text{ ns}$) between low and high light-output levels was demonstrated. The observed bistability was due to saturation of the optical absorption in the passive region of these structures at high injection currents. The MQW laser devices operated in the wavelength region of the excitonic absorption and their bistability was ascribed to the decrease of this absorption at higher light intensities.

[7] D. S. Chemla and D. A. B. Miller, Room-temperature excitonic nonlinear-optical effects in semiconductor quantum-well structures, *J. Opt. Soc. Am. B* 2, 1155-1173, 1985.

[8] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood and C. A. Burrus, Electric field dependence of optical absorption near the band gap of quantum-well structures, *Phys. Rev. B* 32, 1043-1060, 1985.

Chemical modification of surfaces

J. J. Ponjeé and P. N. T. van Velzen

Some of the materials used in the latest products of the electronics industry go through minute but extremely effective surface modifications as a part of the fabrication process.

Outlines of a new technology

In some sectors of technology the surface of the product cannot be allowed to remain in its 'natural' state. The product and the method of producing it must achieve standards of quality that can only be met by adapting the nature of the surface appropriately. The chemical means used for modifying the *adhesion* and *'release'* behaviour of materials by manipulating the chemical structure of the surface will be discussed in this article as an example.

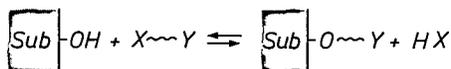


Fig. 1. Diagram illustrating the chemical modification of surfaces. A substrate surface *Sub*, on which reactive groups are present (a hydroxyl group is shown here), is allowed to react with organic molecules possessing two functional groups. One functional group, *X* (red), reacts with a hydroxyl group, the other, *Y* (blue), provides the surface with the required new property.

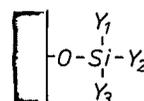
The modification is made by applying a layer of the order of one molecule in thickness to the surface. In general the molecules of the layer must possess two functional groups, one to react with the functional groups of the 'old surface' and one to determine the required new property of the 'new surface' (fig. 1). The bond between these bifunctional molecules and

J. J. Ponjeé is with Philips Research Laboratories, Eindhoven; Dr P. N. T. van Velzen, formerly with these Laboratories, is now with the Philips Industrial & Electro-acoustic Systems Division, Eindhoven.

the surface should preferably be covalent, so that the strength of the new entity is not substantially less than the bonding strength in the substrate. The thickness of the applied layer should have an order of magnitude of no more than one molecular layer, because the bulk properties of the new entity should be relatively close to those of the original substrate. The aim is also to keep the morphology and topography of the surface in the original state.

A precondition for this surface modification is that functional or reactive groups (reactive sites) should be present on the surface, and that they should not be 'shielded' by impurities. This means that the surface must be subjected to a special chemical cleaning procedure^[1] before the actual chemical reaction.

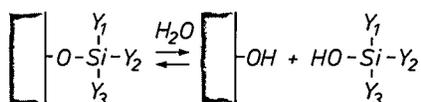
Another precondition for the method is that the interface between the new layer and the substrate should permit stress relaxation on shrinkage and expansion. The bifunctional molecules used for this are generally organic molecules such as organosilanes, which form bonds at the original surface in the following way:



In this structural scheme Y_1 , Y_2 and Y_3 represent organic substituents. It is assumed that the $-\text{O}-\text{Si}-$ or

^[1] A good cleaning procedure is described in J. R. Vig, UV/ozone cleaning of surfaces, IEEE PHP-12, 365-370, 1976.

siloxane bond here can easily be hydrolysed by water, so that in the presence of ambient moisture the following equilibrium is established:



The existence of such an equilibrium (which therefore gives a 'dynamic interface' instead of a rigid network link) would indeed permit the required stress relaxation on shrinkage and expansion.

In the next section we shall first give some examples of the different areas of technology in which this method of chemical surface modification is being successfully used. We conclude that section with a more detailed discussion of one particular area of application: the replication of aspheric lenses, where adhesion and release behaviour are both important.

The application of chemical surface modification obviously requires a thorough knowledge of the surface chemistry. It is necessary to establish whether the reactive sites are indeed present on the surface to be modified, or whether they can be formed there. The degree of surface cleanliness must be known, and the best cleaning procedure. It will also be necessary to know whether the reaction between the reactive groups on the surface and the reagents has in fact been completed in the expected way.

The answers to such questions must be provided by an appropriate analytical method or combination of analytical methods. The following points should be borne in mind when formulating the requirements to be met by such a method.

- Since the chemical modification of the surface is restricted to the outer molecular layer, the number of molecules involved will in general be no more than 10^{13} to 10^{16} per cm^2 . This sets the detection limit (or sensitivity) of the analytical method. Many methods are therefore ruled out from the start.
- Since the majority of reagents used are organochemical in nature, a complete picture of the surface chemistry not only requires data on the *composition* of chemical elements at the surface but also calls for information on the molecular *structure* at the surface.
- The analytical method should be sufficiently 'surface-sensitive' for the results relating to the outer molecular layer to be extracted from the information obtained. This is a particular problem when the bulk material and the surface layer have a comparable structure, which is so for modified polymer surfaces, for example.

We conclude the article with a discussion of two analytical methods that meet these requirements: IETS and SSIMS. Results will be given of some analyses performed with these methods for making aspheric lenses by the replication process.

Examples of applications

Improving the adhesion between a 'novolac'-based photoresist and silicon

The photoresists widely used in IC manufacture have the property of becoming more soluble (positive resist) or less soluble (negative resist) on exposure to illumination. In the subsequent 'development' process the exposed parts of the positive resist (or the unexposed parts of the negative resist) should be easily removable. At the same time the parts of the resist that do not have to be removed should adhere so firmly to the silicon surface that is no danger of them becoming detached during the development.

One of the best known photoresists, which we shall take here as an example, is 'novolac', a group of photosensitive materials used in combination with a phenol-formaldehyde resin [2]. We have found that the use of an aminosilane compound, N-[(3-trimethoxysilyl)propyl]ethylene diamine (*fig. 2*), significantly improves the adhesion between this photoresist and silicon. We assume that this involves three chemical effects. A part (shown in red) of the methoxysilyl groups of the aminosilane molecule reacts with hydroxyl or silanol groups at the silicon surface. These silanol groups always form on the silicon surface in the presence of ambient moisture. Some of the alkaline amino groups (shown in blue) enter into an acid-

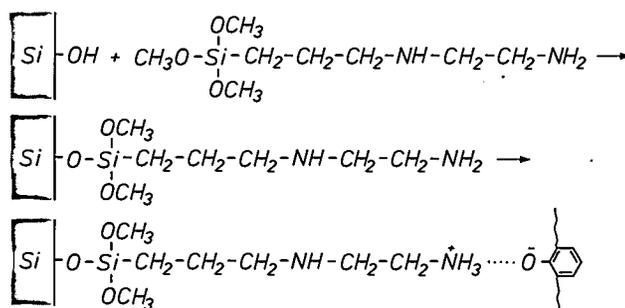


Fig. 2. Reaction equations for the chemical modification of a silicon-oxide surface on which a 'novolac'-based photoresist is to be deposited. The hydroxyl or silanol groups present on the substrate surface form bonds with the aminosilane molecules. The bond is effected by the trimethoxysilyl group of the molecule, shown in red. (The reaction described here relates to a single hydroxyl group with an aminosilane molecule, but two hydroxyl groups may also be involved in this reaction; this also applies to the reactions in figs 3, 4 and 7.) The amino groups, shown blue, are alkaline and can therefore react, in the way shown, with acidic phenolic groups, which are present in the 'novolac' component of certain photoresists.

base reaction with the acidic (mainly phenolic) groups of the novolac. For the resist reaction to occur more than one amino group per silane molecule is required: if only one amino group is present it is highly probable that the interactions of this group with the resin will be inhibited by the silanol groups of the surface, which are also acidic. This effect is sufficiently compensated by the presence of a second amino group.

Improvement of the adhesion between photosensitive polyimide and silicon

In one of the final stages of fabrication integrated circuits are coated with a layer to protect them from α -radiation. A lithographic process is also used in this process step to leave spaces open for the contacts made by the subsequent metallization. A photosensitive polyimide is a suitable material for such a protective layer.

This material is applied in the form of soluble polyamide acids, made insoluble at the appropriate places by illumination through a mask; the complete layer is then converted into a strongly cross-linked polyimide network by curing it at 330 °C [2].

In both the development and curing processes, however, it is difficult to obtain well-defined patterns because there may be a considerable difference in expansion coefficient between the two materials. Another difficulty arising from the difference in expansion coefficient is that the protective coating may come away from the substrate during the curing process. We have achieved good results with the aminosilane compound mentioned above, which takes part in network formation during exposure and curing as indicated schematically in *fig. 3* — i.e. by addition of the amino groups to the unsaturated compound.

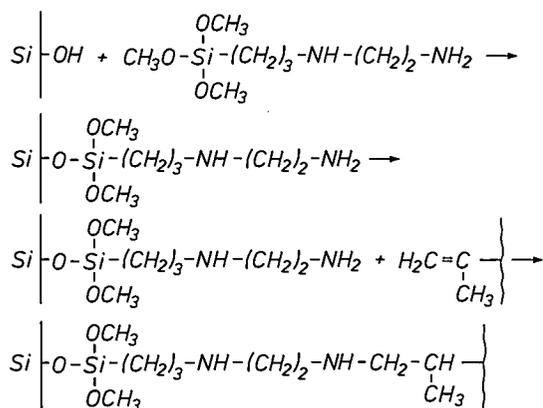


Fig. 3. The same aminosilane compound as in *fig. 2*, which has reacted in the same way with a silanol group of the silicon substrate, and then forms a bond with a polyamide chain (shown here as a wavy line). In the latter case the bond has been formed between an amino group (shown in blue) in the aminosilane compound and an unsaturated bond in the polyamide chain.

Improving the adhesion between noble metals and glass

Certain substrates such as glass can be given desirable reflection properties by depositing noble metals such as gold, silver and platinum on their surfaces. It will often be necessary to preserve the original shape of the substrate surface as far as possible, especially if it is flat. This gives problems if the noble metals do not adhere to the substrate very readily.

The adhesion can be improved by first allowing the substrate to react with mercaptosilane molecules, which contain SH functional groups (*fig. 4*). The adhesion is improved because the noble metals form bonds with the SH groups. The metal can then be grown on the bonded metal atoms by vacuum evaporation, for example.

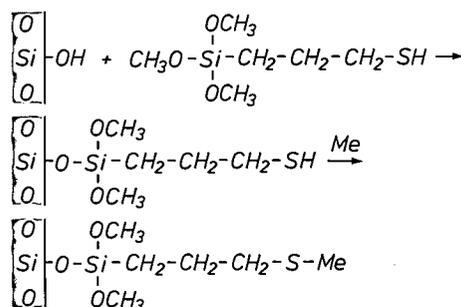


Fig. 4. Noble metals (*Me*) such as gold, silver and platinum can be bonded to glass and other surfaces by a mercaptosilane compound. The metal atoms bonded in this way act as nuclei on which the required metallization layer can be produced in a conventional vapour-deposition process.

Improving the adhesion of the silver mirror coating on video discs

On video discs such as the Philips LaserVision video disc, the silver coating that reflects the laser light must adhere firmly to the substrate material of the disc, which is polymethyl methacrylate (PMMA). This firm adhesion must be achieved without essentially changing the surface pattern of pits that contains the recorded information. These difficult requirements can be met by first applying a monomolecular layer of tannin to the surface, and then depositing tin on the tannin by adsorption from solution.

Just why tannin adheres to the surface of synthetic materials is not as yet entirely clear. It is however known that the tannin forms complexes with Sn(II) ions. These Sn(II) ions adhering to the tannin form silver atoms from silver ions by a redox reaction. The silver atoms deposited then act as nuclei for the 'electroless' deposition of a silver reflecting coating.

[2] L. K. H. van Beek, *Polymer chemistry in the electrical industry*, Philips Tech. Rev. 42, 149-159, 1986.

Adhesion and release in the replication method of making aspheric lenses

As our last example we shall discuss the application of chemical surface modifications in the fabrication of aspheric moulds and lenses by the replication method.

The procedure followed in this method is first to make a roughly *spherical preform* ('body') in glass and then to apply a thin film of resist to it. The resist would typically be a methacrylate resist, which polymerizes on exposure to ultraviolet light. The resist film is cured in the required *aspheric* shape by pressing it against an aspheric mould and then illuminating it (fig. 5). The *aspheric mould* from which the preform is obtained is made with a very-high-precision lathe such as the Colath^[3] which machines a relatively soft material like aluminium extremely accurately to produce the required aspheric shape.

In this replication method there are problems of both adhesion and release. We have solved the problem of *release* from the aluminium mould by allowing the aluminium surface to react with a trimethylsilane

compound, N,N-diethyl-1,1,1-trimethylsilylamine, as shown in fig. 6. This compound reacts with the hydroxyl groups that always form on an aluminium surface exposed to the atmosphere, and in the process the trimethylsilyl groups form covalent bonds with the surface while the amine is desorbed. This reaction is limited to a single monolayer. In this way the aluminium becomes coated with inert hydrophobic groups that completely inhibit adhesion to the surface and therefore facilitate release.

To improve the *adhesion* of the photopolymerized resist to the preformed spherical bodies in glass, the surface of the glass can be modified in the way shown in fig. 7. The functional groups in blue, the methacrylate groups, are of a similar nature to the functional groups that initiate and take part in the network formation on exposure to illumination.

Analysis of chemically modified surfaces

In the introduction we mentioned the special requirements placed on the analytical methods to be used in this work. In view of these requirements we have mainly used inelastic electron-tunnelling spectroscopy (IETS) and static secondary-ion mass spectroscopy (SSIMS) for the applications considered here.

IETS; the method^[4]

Inelastic electron-tunnelling spectroscopy is a method that can be used to measure the vibrational spectra of molecules adsorbed on a metal-oxide surface. The method is based on the tunnelling of electrons through an insulator. Fig. 8 is a diagram of the cross-section of a tunnel diode used in this method.

It is known from quantum mechanics that in a structure like the one in fig. 8 the probability that electrons from metal 3 will be found in metal 4 at a particular time is not zero for very thin insulators but has a finite value — even if the energy of the electrons in that metal is lower than the energy barrier formed by the insulator. The converse is also true; as long as there is no potential difference between 3 and 4, the net tunnel current is zero. If a voltage is applied, a tunnel current will flow. By analogy with conventional conductors, a tunnel resistance can then be defined that in theory will increase exponentially with the thickness of the insulator. The resultant tunnel current consists of two components (fig. 9): an 'elastic' tunnel current (1) that gradually increases as a function of the applied voltage (from $V = 0$), and an 'inelastic' tunnel current (2), which also increases gradually, but only from a threshold value E_{\min} . This threshold value E_{\min} is related to the energy transferred by the electrons to the molecules or the atoms

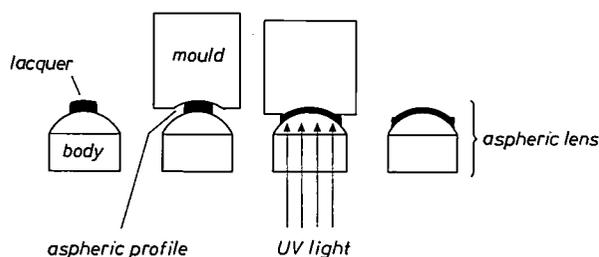


Fig. 5. The principal process steps in the replication method for making aspheric moulds and lenses.

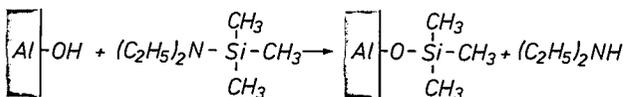


Fig. 6. Reaction scheme for producing a layer with good release properties on an aluminium substrate. The trimethylsilyl groups (in blue), which ultimately constitute the substrate surface, provide no opportunity for adhesion to the surface.

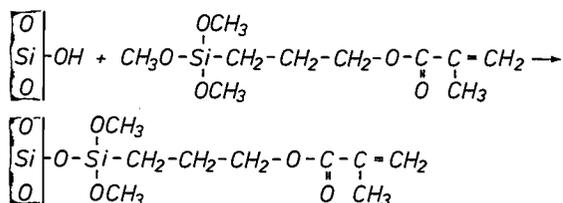


Fig. 7. Reaction equation for a layer to improve the adhesion to a glass substrate. The terminal unsaturated bond (blue) is highly reactive and participates in the network formation that occurs as soon as a photopolymerizing material on this layer is exposed to illumination.

(or both) of the insulator. The frequency ω of the excitations generated by the electrons can be found from $E_{\min} = eV_{\min} = h\omega/2\pi$, where h is Planck's constant and e the electronic charge.

In reality, of course, a layer of organic molecules on an insulator would not have a single threshold value E_{\min} but several, corresponding in theory to the existing number of molecular vibrations and electronic excitations.

The region we are most interested in is the one corresponding to a potential difference of between 40 and 500 meV, which contains most of the molecular vibrations (with wave numbers between 300 and 4000 cm^{-1}).

Accurate measurement of the current-voltage characteristic of a tunnel diode, as in fig. 8, will thus give the vibrational spectrum of the organic layer

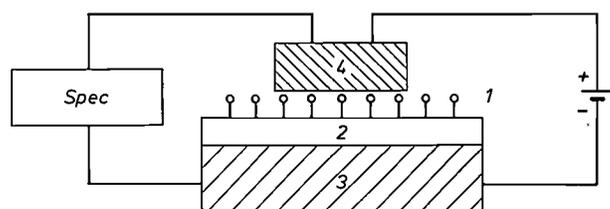


Fig. 8. Diagram of the cross-section of an arrangement for measuring IETS spectra. The organic molecules 1 under investigation are applied to an insulating layer 2 sandwiched between two metal layers (3 and 4); the complete device forms a tunnel diode. Spec IETS spectrometer. See also fig. 9.

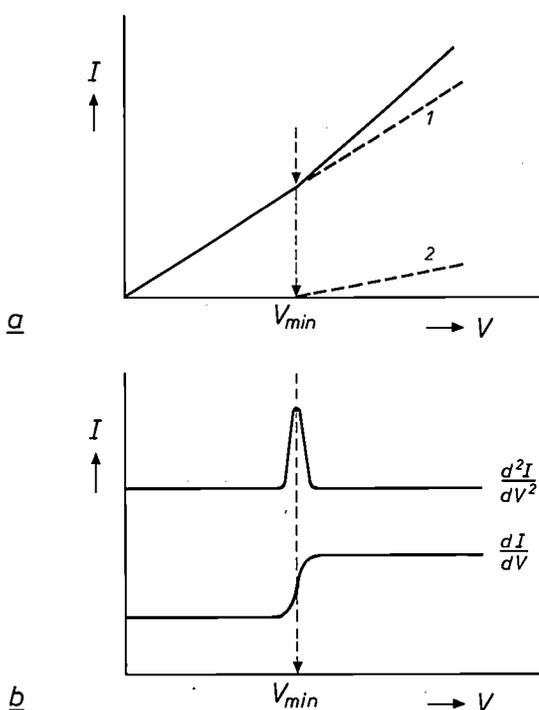


Fig. 9. a) The total tunnel current I generated as a function of a voltage V in the arrangement in fig. 8 consists of an 'elastic' component (1) and an 'inelastic' component (2), occurring after a threshold value V_{\min} . b) In practice, modulation techniques are used to determine the second derivative d^2I/dV^2 directly as a function of V , which is proportional to d^2I/dV^2 .

on the surface of the insulator. It turns out, however, that the change in the tunnel current as a result of 'opening' an inelastic tunnel channel (the increase in the slope at V_{\min} in fig. 9a) remains well below 1%, and is almost undetectable. These threshold values can nevertheless be determined as peaks at the potential difference V_{\min} by measuring the tunnel current as a function of the second derivative of the potential difference, as illustrated schematically in fig. 9b.

At this point we should say something about the potential of the IETS method — and its limitations.

- The detection sensitivity of IETS is far below one monomolecular layer.

- IETS measurements are made at 4.2 K or below to reduce thermal line-broadening. The top metal (4) generally used is lead, which is a superconductor at this temperature, thus slightly improving the resolution. In practice, linewidths of 15-25 cm^{-1} are achieved, and this is amply sufficient to discriminate between the relevant peaks, even for complex organic molecules. While lead has the advantage that it is a superconductor at 4.2 K, it can also be evaporated under relatively mild conditions, and it is a metal that is fairly inert to organic compounds.

- A limitation of the IETS technique is that it can only be used on metals on which a very thin (about 2 nm) and electrically stable insulator can be formed reproducibly. Because of the very strict conditions that apply to the breakdown voltage of the insulating material ($> 2.5 \times 10^6 \text{ Vcm}^{-1}$), tunnel spectroscopy has so far been limited to aluminium, magnesium, tin and lead. Most experiments have been performed on aluminium-oxide surfaces.

IETS; analysis results

As an example of the application of IETS for the analysis of chemical surface modification in practical situations, we shall discuss the modification of aspheric aluminium moulds (which have been oxidized at the surface) with N,N-diethyl-1,1,1-trimethylsilylamine. Because the quality of the insulating oxide layer for IETS measurements must be extremely high it is impossible to perform measurements directly on the mould. The measurements are therefore carried out on aluminium substrates vapour-deposited on glass.

Fig. 10a gives the IETS spectrum of an uncoated surface of aluminium oxide. Only two bands are really distinct and easily assigned; these are the band at 940 cm^{-1} , due to an Al-O 'stretching' vibration, and

[3] T. G. Gijsbers, COLATH, a numerically controlled lathe for very high precision, Philips Tech. Rev. 39, 229-244, 1980.

[4] P. K. Hansma, J. Electron Spectrosc. & Relat. Phenom. 30, 163-174, 1983.

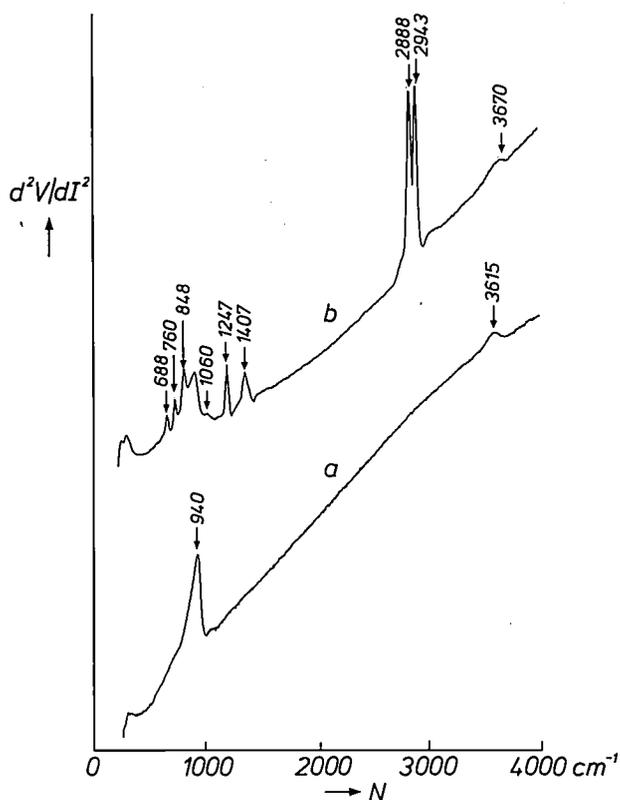


Fig. 10. IETS spectra of (a) a clean aluminium-oxide surface and (b) the same surface after reacting with N,N -diethyltrimethylsilylamine. The values determined for d^2V/dI^2 are given in arbitrary units as a function of the wave number $N (= \omega/2\pi c)$, where c is the velocity of light. See the text for the identification of the peaks in the spectra, and fig. 6 for the reaction equation.

the band at about 3615 cm^{-1} , due to the O-H stretching vibration of hydroxyl groups at the surface. The initial requirements for the chemical surface modification, as expressed above — a clean surface with reactive sites, in this case the aluminium-hydroxide groups — have therefore been met.

After this surface was brought into contact with the vapour of the aminosilane mentioned above, the IETS spectrum of the resultant aluminium-oxide surface was measured (fig. 10b). It is clear that the new surface no longer consists of clean aluminium oxide. The complete spectrum can be interpreted with the aid of infrared and Raman data for vibration frequencies of functional groups in organic molecules. This will not be discussed here in detail [6].

The main conclusions that can be drawn from the measurements on this system are mentioned below. Most of the bands in fig. 10b can be directly assigned to the $\text{Si}(\text{CH}_3)_3$ group, and no indications can be found in the spectrum for the presence of a functional $\text{N}(\text{C}_2\text{H}_5)_2$ group. The bonding of these $\text{Si}(\text{CH}_3)_3$ groups to the aluminium oxide can also be derived from a band (of low intensity) at 1060 cm^{-1} ; this band can be assigned to the asymmetric Al-O-Si stretching

vibration in the $\text{Al-O-Si}(\text{CH}_3)_3$ structure. It follows that the model for this surface reaction presented in fig. 6 is in good agreement with the results obtained with IETS.

The shape of the spectrum and the intensities of the bands do not depend on the length of time the oxide is exposed to the aminosilane vapour, which indicates that the reaction is fast and also, in complete agreement with the model, is limited to the formation of one monomolecular layer. Although in ordinary conditions the Al-O-Si bond can easily be broken in the presence of water (see the introduction), under these conditions it takes more effort to remove the $\text{Si}(\text{CH}_3)_3$ groups from the surface with water. This is because the new surface is made hydrophobic by covering it with these groups, and cannot therefore be wetted easily. The surface layer thus obtained therefore possesses excellent release properties. Although these measurements were performed on a model system, there is no reason to assume that a thoroughly cleaned aluminium mould (on which a clean oxide is present) cannot be modified in an identical manner.

SIMS; the method [6]

The outer molecular layer of a material can be analysed with SIMS. The surface of the material is bombarded with energetic particles (energy 1-20 kV), generally ions but atoms are sometimes used, and the bombardment releases secondary particles from the surface. This is the process known as sputtering. A small proportion of the sputtered particles are secondary ions — both positive and negative — which can be detected by the principles of mass spectrometry. Depending on the energy and current density of the primary ion beam, there are two limiting cases:

Dynamic SIMS. In this case relatively high beam-current densities are used. The molecular or atomic layers of the material are sputtered at a high velocity mainly as elemental ions. Dynamic SIMS is particularly suitable for determining the *elemental composition* of surfaces (e.g. the depth profile of the doping in semiconductor material); the method cannot generally provide information about the chemical *structure* of the molecules at the surface.

Static SIMS. When low-energy primary ions are used at a low current density (six orders of magnitude less than in the dynamic case) during sputtering, cluster ions of higher mass originating from the outermost monolayers are observed. The structure of ions of this kind provides information about the chemical struc-

[5] P. N. T. van Velzen and M. C. Raas, *Surf. Sci.* **161**, L 605-L 613, 1985.

[6] A. Brown and J. C. Vickerman, *Surf. & Interface Anal.* **6**, 1-14, 1984.

ture of the surface (see the introduction). Under these 'static' conditions the total amount of material is small, so that the mass spectrum of the outermost layers of the surface can be measured quasi-non-destructively. This implies that only a negligible fraction of the surface is damaged during the entire measurement (usually about 10^{-3} of a monolayer or less), while the probability that a point on the surface is struck more than once by a primary ion is very small. For the measurement, however, this requires highly sensitive mass-spectroscopic detection of the secondary ions. This is achieved by means of time-of-flight analysis of the secondary ions. This explains the name of the method: time of flight SIMS, or TOF-SIMS; see fig. 11.

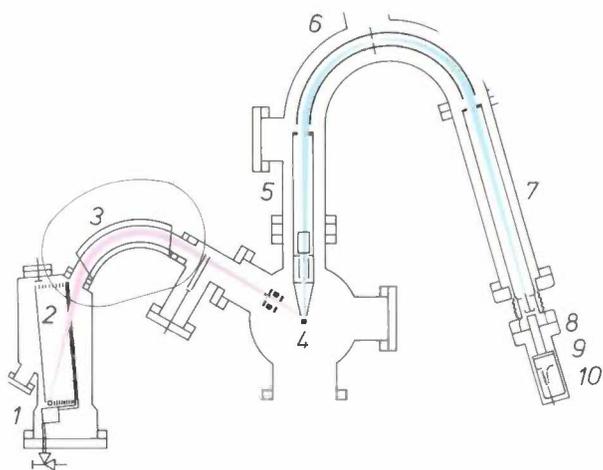


Fig. 11. Diagram of the TOF-SIMS spectrometer used for measuring the SIMS spectra given here. It is located in Prof. A. Benninghoven's Surface Physics group at the University of Münster, West Germany. The instrument consists of a primary ion source (1), a deflection plate (2), a magnet for mass discrimination and focusing the primary ion beam (3), a multiple-focus time-of-flight analyser (5, 6, 7) and a detector (8 to 10). In this instrument a pulsed mass-discriminated primary ion beam is generated (pulse width 10 ns) with an accurately defined energy and a very low mean current (of the order of 1 pA), which sputters secondary ions from the specimen (4). These ions are extracted from the surface and accelerated by a pulsed voltage on the entrance plate of the linear drift space (5). Mass discrimination of the secondary ions is obtained because of the differences in time of flight through the combination of two linear drift spaces (5, 7) and an electrostatic sector field (6); the time of flight is proportional to \sqrt{m} , where m is the mass of the ions.

The ions are detected by a single channel plate (8), and a scintillator (9) optically linked to a photomultiplier (10). The detector operates in the single-ion counting mode.

As compared with other methods of secondary-ion mass spectroscopy in the static limit range (e.g. quadrupole-static SIMS), TOF-SIMS has the following very important features:

1. High transmission of the mass-discriminating system ($> 10\%$).
2. Quasi-simultaneous measurement of all the ions of the same sign (the entire mass range is measured for each primary-ion pulse).
3. Good mass discrimination ($m/\Delta m \approx 1000$, where m is the mass and Δm the peak width at half height).
4. Because of 1 and 2 the sensitivity of the spectrometer is very high. A detection limit of 10^{-6} monolayer or better has been demonstrated for a number of systems.
5. The typical measurement time per spectrum is less than 1 minute.

SIMS; analysis results

We conclude with some results of analyses relating to improvement in the adhesion of a methacrylate resist on lanthanum-containing glass, as used in the replication process.

The glass is cleaned with an alkaline soap, rinsed with water and then dried in isopropanol vapour. The modification is performed by exposing the surface to the vapour of 3-methacryloxypropyltrimethoxysilane; it is assumed that a reaction as shown in fig. 7 takes place during this exposure. Figs 12 and 13 give the

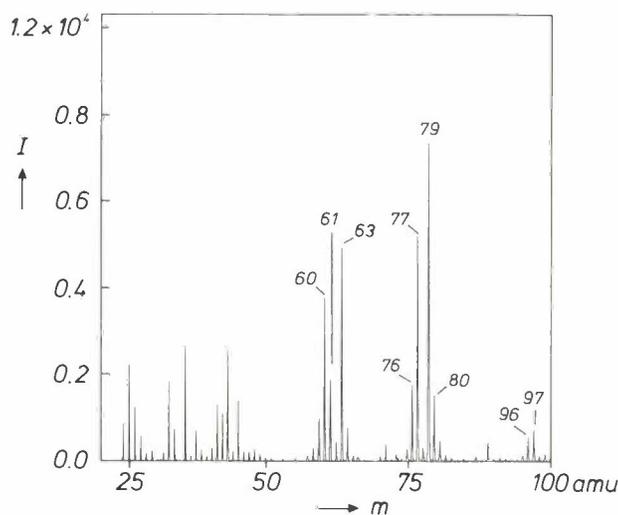


Fig. 12. Negative SIMS spectrum of a cleaned surface of lanthanum-containing glass. Owing to adsorbed residues of alkaline soaps used in cleaning, the spectrum contains peaks attributable to SO_4^- , H_2PO_4^- , PO_3^- , SO_3^- and PO_2^- (at 96, 97, 79, 80 and 63 amu). The large peaks at 60, 61, 76 and 77 amu (for example) correspond to SiO_2^- , SiO_2H^- , SiO_3^- and SiO_3H^- respectively.

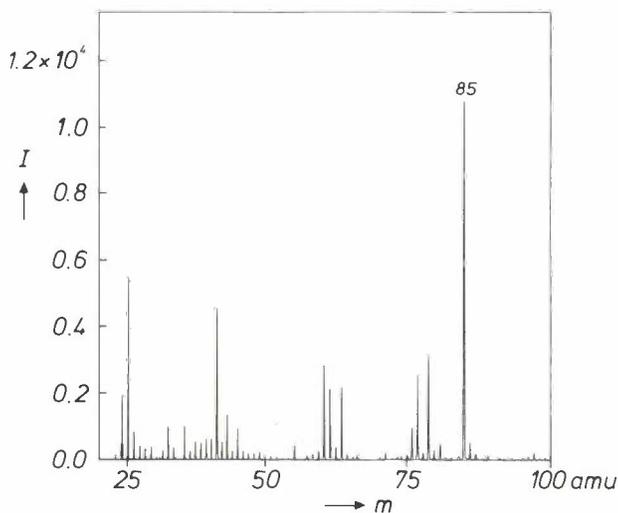


Fig. 13. Negative SIMS spectrum of the same surface as in fig. 12, after it has been allowed to react with a methacryloxysilane. Fig. 7 gives the assumed chemical structure of the modified surface. The largest peak in the spectrum is now observed at 85 amu, and is assigned to the methacrylate anion, $\text{CH}_2=\text{C}(\text{CH}_3)\text{COO}^-$.

negative SIMS spectra of the cleaned and modified glass surface, respectively. Some details of the spectra will now be discussed.

The spectrum of fig. 12 shows that even after cleaning the glass surface is not completely clean. For example, peaks are visible at 96, 97, 79, 80 and 63 amu^[7], which must be assigned to SO_4^- , H_2PO_4^- , PO_3^- , SO_3^- and PO_2^- respectively. These ions originate during the sputtering process from the molecules of residues of the alkaline soaps used for cleaning. The quantities of these soap residues are small, however, as can be seen from a comparison of the intensities of these peaks with those of the intense peaks that originate from the glass substrate, e.g. at 60, 61, 76, 77 amu, and are assigned to SiO_2^- , SiO_2H^- , SiO_3^- and SiO_3H^- .

The same conclusion may be drawn from the positive SIMS spectra of the glass, not shown here, in which the metal ions (including Na^+ and La^+) in the glass can also be observed, demonstrating that these are also present at the surface.

Fig. 13 shows that exposure of the cleaned glass to the vapour of the 3-methacryloxypropyltrimethoxysi-

lane does in fact modify the chemical structure of the surface. The largest peak in the negative SIMS spectrum in this mass range is observed at 85 amu, and originates from the methacrylate anion $\text{CH}_2=\text{C}(\text{CH}_3)\text{OO}^-$, formed by fragmentation of the functional end-group of the methacryloxysilane mentioned above. The reaction with the methacryloxysilane therefore does take place in the way that appears to be necessary for a substantial improvement in the adhesion.

Summary. Bifunctional organic molecules that form bonds with a surface with one functional group give new properties to the surface with the other functional group. Adhesion to a surface or release from a surface can be improved by using this method. Pre-conditions for such chemical surface modification are the presence of reactive sites at the surface, the absence (because of good cleaning procedures) of impurities at these sites, and the possibility of stress relaxation on expansion and shrinkage. Subjects discussed include improvement of the adhesion between a photosensitive resist based on 'novolac' and silicon with an aminosilane, improved adhesion between noble metals and glass with a mercaptosilane, improved adhesion of silver reflecting coatings on LaserVision discs with tannin, an improvement of both adhesion and release in the replication method used for making aspheric lenses, using a methacryloxysilane and a trialkylsilane. The chemical surface modifications described here for the replication method have been investigated by IETS and TOF-SIMS. The principles of these methods are discussed and some results are presented.

^[7] (Unified) atomic mass unit.

Striations in a gas discharge

F. C. van den Heuvel

Optical instabilities in gas discharges have been known almost as long as gas discharges themselves. Fortunately, in most applications their occurrence can be avoided, often by taking external measures adapted to a particular situation. However, a universal remedy requires a better understanding of these effects.

Introduction

In the first half of the 19th century it was discovered that a gas between two electrodes in a sealed glass tube can conduct an electric current. It was also found that the gas emitted radiation. This can be considered as the start of the application of these 'gas discharges' for artificial lighting, since the emitted radiation can have a wavelength in the visible range, or fluorescence from materials on the glass wall can convert the radiation into visible light. In the earliest days strange optical effects were occasionally observed: a pattern of alternate bright and dark regions appeared along the axis of the gaseous column (see *fig. 1*). These discharges are said to contain 'striations', or are described as being 'striated', terms taken from the Latin word 'striae', meaning 'stripes'.

In modern artificial lighting low-pressure mercury/inert-gas discharge lamps (TL, PL and SL lamps) play a leading part. In the design of these lamps efforts are made to create conditions in which the optical effects mentioned above either do not occur or cannot be observed with the naked eye. An artificial light source is obviously required to emit its radiation as uniformly as possible. The parameters that affect the formation of striations in the gas discharge — the composition and pressure of the gas mixture, the dimensions of the tube, the discharge current and the mercury vapour pressure — should be controlled so as to prevent the occurrence of striations, without causing any direct conflict with other characteristics required of the lamp. It is important, for example, that the lamp converts electrical energy into light efficiently, has a long life and is compact. For standard interior lighting it is usually fairly

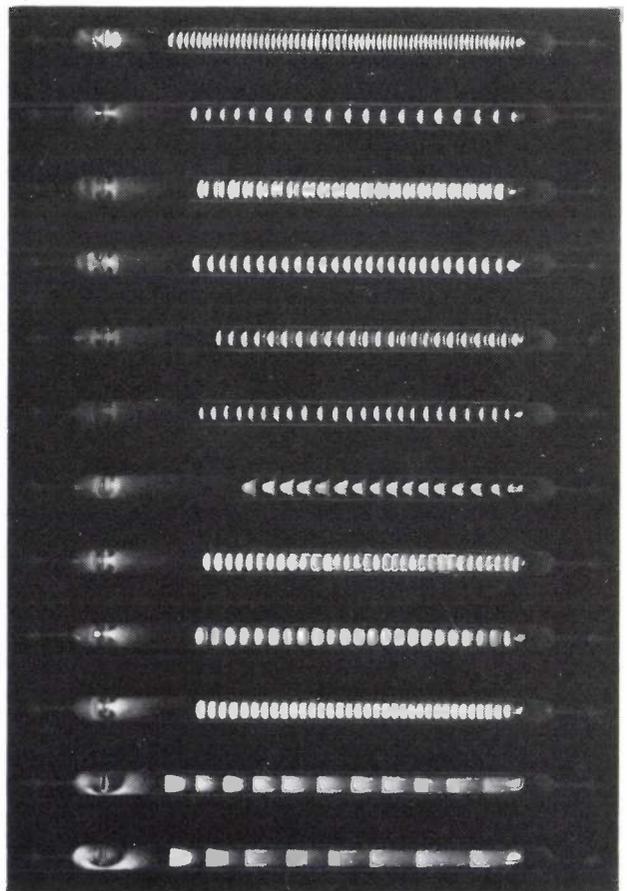


Fig. 1. Some examples of striations in gas discharges. This photograph dates from the early days of gas discharges^[1] (1878!). The striations that can be seen in the fourth discharge from below are the ones that most closely resemble an 'instantaneous exposure' of the striations seen in our mercury/argon discharges.

[1] This photograph has been taken from: S. C. Brown, A short history of gaseous electronics, in: M. N. Hirsch and H. J. Oskam (eds), Gaseous electronics, Vol. 1, Electrical discharges, Academic Press, New York 1978, pp. 1-18.

easy to find a compromise, because the (r.m.s.) discharge current and the operating temperature, which determine the mercury vapour pressure in the tube, are not subject to marked changes. Even if a compromise should prove difficult it is often still possible to prevent the appearance of striations by taking appropriate external measures. For example, the striations in a fluorescent lamp that fails to reach the required operating temperature because the ambient temperature is too low can be avoided by enclosing the lamp in a suitably modified fitting. Such an approach will not work, however, when a lamp is required to operate without striations in varying conditions. This is the case, for instance, with the integrated lighting systems nowadays installed in large buildings because they are economical of energy. In these systems the lighting level is controlled from a central point to complement the amount of daylight entering the building. There is then a danger that the discharge current through the lamps will become so low that striations appear. The effect becomes worse because a low discharge current gives a lower operating temperature and hence a lower mercury vapour pressure. New precautions then have to be taken to prevent this from happening.

This article is concerned with the way in which the electrical circuit to which the gas discharge is connected affects the occurrence of striations. The discussion will be confined to a d.c. discharge with the current kept just below the critical value at which striations appear. Before considering our experiments and the results, let us first briefly look at the gaseous discharge, the occurrence of striations and an explanation of how they come about. Later in the article there is a discussion of our observations of striated discharges and their relation to the electrical circuit. It will be shown that the striations excite an oscillation in the electrical circuit, and that this oscillation maintains them.

Striations in gas discharges

Fig. 2 gives a diagram of the spatial distribution of an unstriated gas discharge. The distribution is based on the intensity of the generated radiation. Most of the discharge between the two electrodes is taken up by the positive column, which is primarily responsible for the radiation output from the discharge. The phenomena we are concerned with, the striations, appear in this column.

The concentrations of electrons and ions in the positive column are practically identical: macroscopically, the plasma is electrically neutral. The temperature of the electron gas, which is a measure of the velocity of the electrons, is relatively high (about

12 000 K), whereas the temperature of the ion gas is only a few hundred K. The highly mobile electrons with a drift velocity of about 10^4 m/s are responsible for the electric current. The contribution from the much slower ions is negligible. This difference in drift velocity of electrons and ions might have the effect of endangering the neutrality of the plasma, for if the concentration of electrons (and hence ions) is higher in a particular zone of the discharge, the electrons will disappear from this region faster than the ions. This situation is not reached, however. With a separation of even as little as $10\ \mu\text{m}$ between the ion and electron distributions such a strong opposing field is formed that any further drifting apart is prevented. In this way the plasma opposes any large-scale build-up of space charge^[2]. This type of field is known as 'ambipolar'. It reduces the applied axial field where the electron density is relatively high and increases it where the electron density is low. This happens in such a way that there is no spatial variation in the electric current: $\text{div } j = 0$.

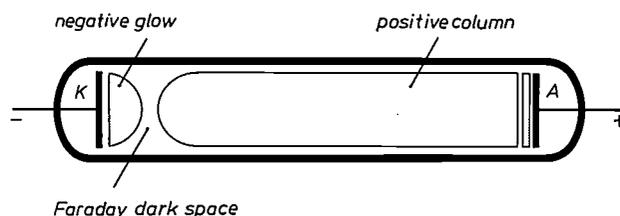


Fig. 2. Spatial distribution of a gas discharge on the basis of the variation in the intensity of the emitted light. Only the most important parts of the discharge are indicated. *K* cathode, *A* anode.

Under certain conditions striations can occur in a gas discharge. Although the effect was known in the early days of gas discharges, no explanation for it appeared till about thirty years ago, when L. Pekarek^[3] put forward the following argument. Suppose that, for one reason or another, an unstriated positive column contains a region of some magnitude ($\gg 10\ \mu\text{m}$) where the ion density (and hence the electron density) is higher than in the rest of the column. Electrons drifting through this region are then subject to a reduced field because of the presence of the opposing field described above. The electron gas therefore becomes cooler, so that fewer ions are produced by ionization. Since the cooling and heating of the electron gas by an electric field are not instantaneous, of course, the region with the lower temperature will be displaced with respect to the region with the higher density. It will take about $1\ \mu\text{s}$ before the field starts to affect the temperature. In that time the drifting electron gas will

have covered a distance of no less than a centimetre. A local increase in the ion density therefore gives a corresponding decrease in ion density in the adjoining region, and in turn this leads in due course to an increase in the next region, and so on. In the plasma the local temperature profile and the diffusion determine the dynamics of the ion and electron density profiles. Assuming a uniform positive column with a local change in the ion and electron density, Pekarek showed that the assumed local density distribution gives rise to the formation of a kind of 'wave packet'. This propagates along the axis of the discharge, expanding at the same time. The process becomes visible because the intensity of the radiation field changes, as a result of the changed electron temperature. This was how Pekarek explained the origin of moving bright and dark regions in a striated gas discharge.

Depending on the conditions in the plasma, one of the following situations may now arise. If the amplitude of the waves in the packet decreases and the plasma returns to a situation with no perturbations, the plasma is then by definition stable. However, the conditions in the plasma may be such that the amplitude of the waves increases. If the packet then moves more slowly than it expands, the plasma becomes completely unstable. In such a case the plasma does not return to the state of equilibrium. Between these two extremes the plasma is convectively unstable. The term 'convectively' means that the amplitude of the perturbation increases and that its displacement is faster than the spatial growth. The plasma is locally unstable while the packet is passing by, but afterwards a local equilibrium sets in again. We have studied the occurrence of striations in a convectively unstable plasma of this type.

The effect of the electrical circuit on the striations

Fig. 3 shows the circuit used for the experiments. The operating temperature is kept artificially constant by surrounding the discharge tube with a second glass tube and circulating water at a constant temperature of 50 °C in the space between them. The part of the circuit that supplies the direct current for the discharge is decoupled from the load circuit — an a.c. circuit that contains a variable resistance — by an inductance and a capacitance. These have values such that their influence on the load is negligible. The effect of the resistance on the occurrence of striations in the discharge was investigated. The current through the discharge was set just below the critical value; this made the plasma convectively unstable.

The 'wave-packet' perturbations described by Pekarek were indeed observed. They arise as a conse-

quence of noise in the neighbourhood of the cathode and travel towards the anode at a velocity (group velocity) of about 60 m/s. The intensity of the complete wave packet increases as it moves away from the cathode. The phase velocity of the waves 'contained' in a packet has the opposite direction.

The electrical circuit has a definite influence on the effects in the discharge, as can be seen in fig. 4. As the external resistance is reduced, the amplitude and the spatial magnitude of the packets increase. When the resistance is sufficiently reduced, wave packets overlap to such an extent that they cannot be distinguished separately and combine to form a travelling wave with a more or less constant wavelength and frequency. It seems as if the wave travels in the opposite direction to the wave packets, but this is only because the phase velocity is positive towards the cathode.

The amplitude of the wave was measured with a moveable photodiode as a function of the distance to the cathode; see fig. 5. The photodiode signal consists of two contributions: a signal that would also be measured in an unstriated gas discharge (with the same value everywhere along the column) and a small variation due to the occurrence of striations. The figure only shows the contribution due to the striations. At every position only the maximum value of the variations in the light intensity was recorded. The

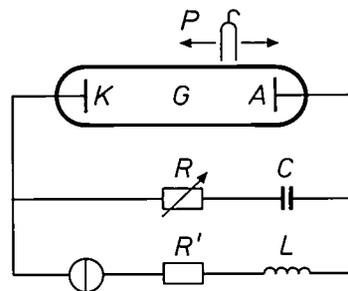


Fig. 3. The experimental arrangement. The discharge tube contains a mixture of argon and saturated mercury vapour at a total pressure of 400 Pa. The tube is 0.6 m long and has an inside diameter of 12 mm. The distance between the electrodes is 0.4 m. The gas discharge *G* is incorporated in an electrical circuit. The effect of the variable resistance *R* on the phenomena in the discharge was investigated by measuring the intensity of the emitted radiation with a photodiode *P*. The capacitance *C* and the inductance *L* are included in the circuit to decouple the direct-current and alternating-current parts of the circuit.

[2] The field reaches equilibrium very quickly. At a drift velocity of 10^4 m/s a charge separation of 10 μm is reached after only 1 ns. The charge separation in the plasma is comparable with a dielectric between two capacitor plates. A charge separation on a microscopic scale leads to a macroscopic field, with the dielectric as a whole remaining electrically neutral.

[3] L. Pekarek, The development of a pulse-disturbance in a dc-plasma, Proc. 6th Int. Conf. Phen. Ion. Gases, P. Hubert and E. Cremieu-Alcan (eds), Vol. 2, Paris, 1963, p. 133.

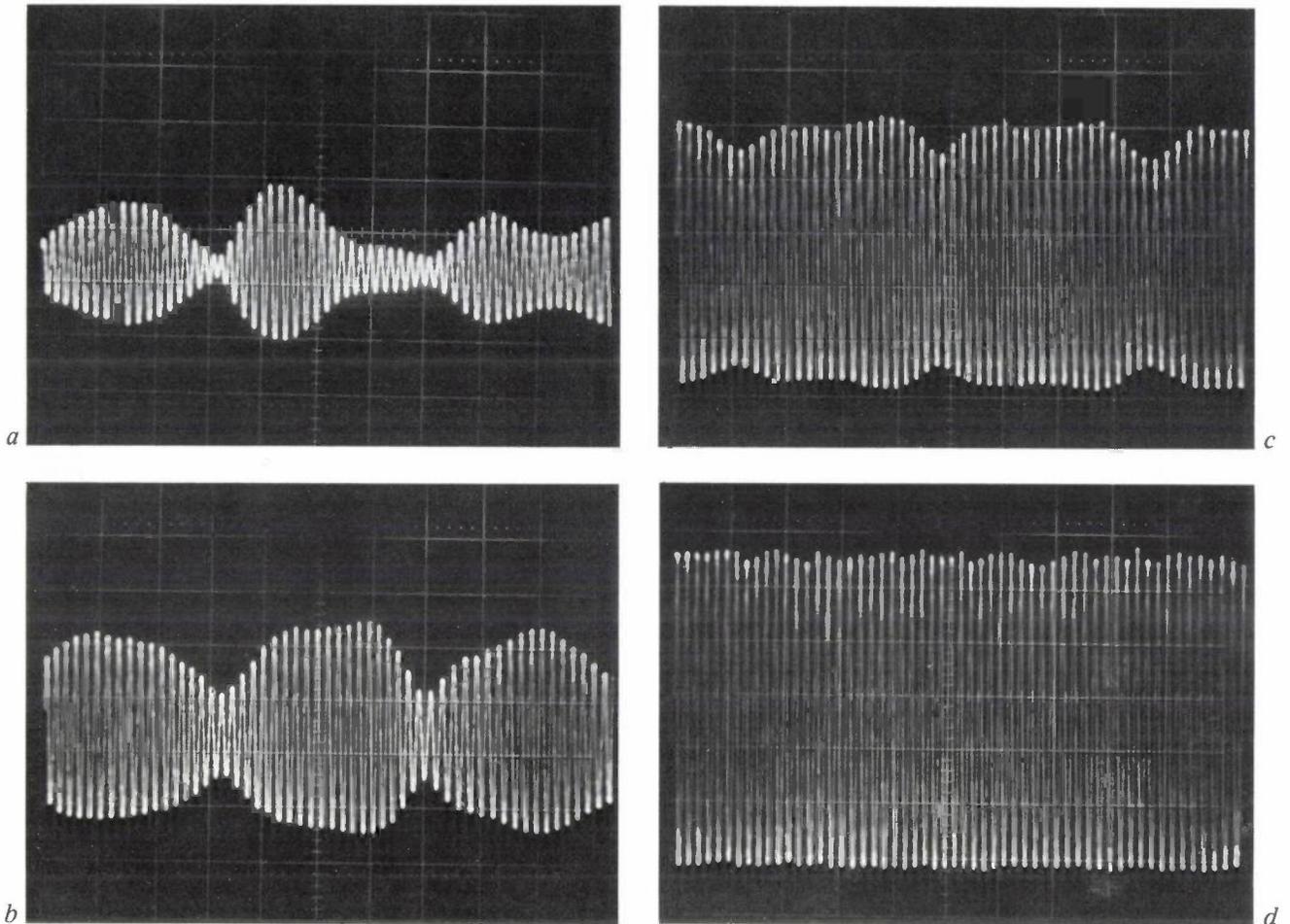


Fig. 4. Variation in the intensity of the emitted radiation, ΔI , as a function of time, measured at a position close to the anode for four values of the variable resistance R . a) $R = 6 \text{ k}\Omega$, b) $R = 3.7 \text{ k}\Omega$, c) $R = 1.9 \text{ k}\Omega$, d) $R = 0.2 \text{ k}\Omega$. The size and amplitude of the wave packets increase as the resistance decreases, until finally a travelling wave is formed. Similar pictures taken at the cathode are qualitatively identical, the only difference being that the size and the intensity of the wave packets are respectively smaller and lower.

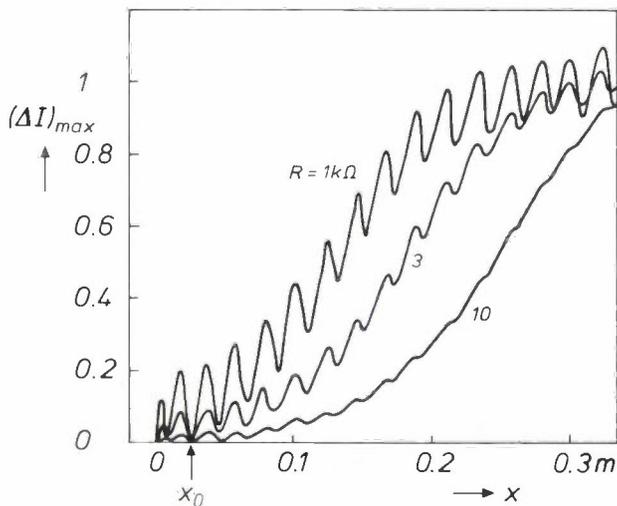


Fig. 5. The maximum value $(\Delta I)_{\max}$ of ΔI (the variation in the radiation intensity) due to the occurrence of striations, measured with the photodiode, as a function of the distance x to the cathode. The constant contribution to the photodiode signal has been omitted. At the position x_0 the value of $(\Delta I)_{\max}$ is always zero.

actual travelling wave is therefore not shown in the figure, but the variation in the amplitude of this wave as a function of the distance to the cathode. Apart from the sinusoidal modulation — which will be discussed later — it can be seen that the spatial amplitude of the wave increases exponentially with the distance from the cathode, as does the intensity of the individual wave packets. At the anode the wave amplitude reaches a saturation value. When R is decreased the intensity and stability of the wave increase, and the saturation value is reached before the anode. Closer inspection of the curves shows that the exponential growth is identical for the three curves. The rate at which the spatial amplitude of the wave increases is not affected by the resistance and is therefore a property of the plasma. The sinusoidal modulation of the wave amplitude is constant along the column and, like the intensity of the wave, it increases with decreasing R . Fig. 5 also shows that the total wave am-

plitude is zero a few centimetres away from the cathode. We shall return to this point presently.

The most important finding is that the electrical circuit has the effect that the convectively unstable plasma no longer returns to the unstriated situation, which means that the travelling wave is obviously in a state of continuous self-excitation. Because the electrical circuit is known to have an effect on the wave phenomena, the obvious course is to treat the discharge and the circuit as a single entity. The wave is maintained by feedback via the circuit. This situation is rather like that of a noise amplifier that has sufficient feedback to bring it into saturation so that it becomes a self-stabilizing oscillator. The essential difference, however, is in the way in which the coupling between the striation effects in the tube and the circuit is obtained. In the circuit, information can only be transferred in the form of electric currents. The occurrence of a stabilized travelling wave in the discharge means that an alternating current at the same frequency as that of the striation wave must flow in the circuit.

In the next section it will be shown, from analyses of the photodiode signal and the current in the external resistance, that this qualitative assumption can be quantitatively justified.

The oscillation in the circuit

The existence of the oscillation, which is assumed to be responsible for initiating and maintaining the travelling wave in the discharge, can be demonstrated in a number of ways. First of all it will be shown that characteristics of striations can be found both in the photodiode signal and in the electric current. An explanation for this will then be derived from a mathematical description of the wave phenomena.

Measurements

The observations described in the previous section lead to the assumption that the perturbations in the discharge maintain themselves by feedback in the electrical circuit. To test the validity of this assumption, we made a closer analysis of the signals passing through the discharge and through the electrical circuit. A spectrum analyser was used to record the frequency content of the photodiode signal; see *fig. 6*. These spectra were taken for a gas discharge whose discharge current was a little closer to the critical current than in *fig. 5*. This implies that growing and overlapping wave packets only barely reach the saturated amplitude at the anode (small R) or do not reach it at all (large R). When the resistance is high ($R = 4 \text{ k}\Omega$) the diode signal is irregular, with a noise-gain band

around 2.7 kHz. When the feedback is increased (by reducing the resistance) this band remains visible, but increasingly sharper resonance peaks also appear. This picture resembles the frequency spectrum of a laser. Spectra taken of the voltage across the discharge or — and this amounts to the same thing — of the

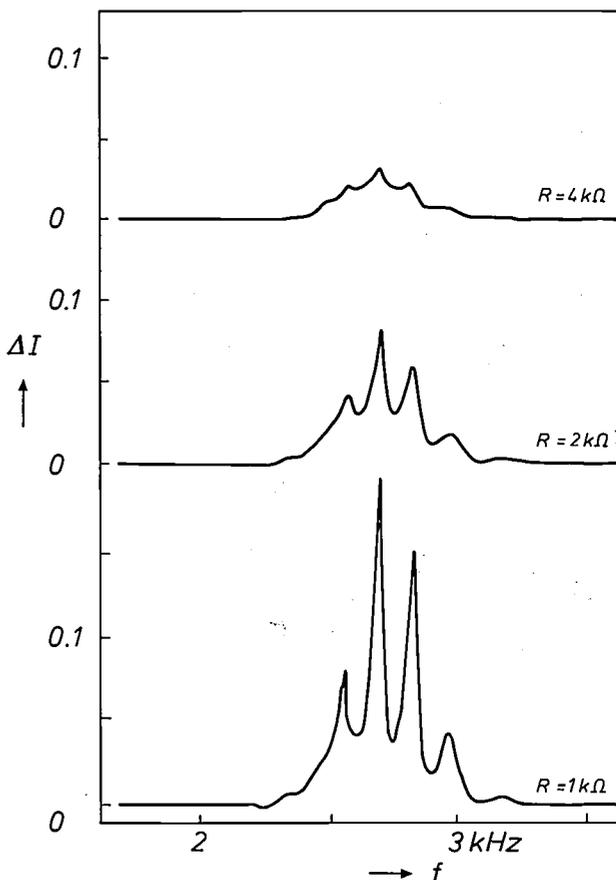


Fig. 6. Frequency spectrum of the photodiode signal for three values of the resistance R . At smaller values of R resonances occur.

current in the feedback resistance, show the same picture; the maxima are at the same frequencies, as can be seen in *fig. 7*.

So the alternating currents assumed do indeed occur, although they are very small — a few tens of μA as compared with the direct current of 100 mA. It appears that the striated discharge acts as a kind of alternating voltage source. In Pekarek's model of a striated discharge, striation waves are accompanied by a local field $E(x, t)$ that ensures the continuity of the current ($\text{div } j = 0$). The electric field also takes the form of a travelling wave with an exponentially increasing amplitude. It therefore seems reasonable to extend the model to a voltage source, since the potential difference between anode and cathode is equal to

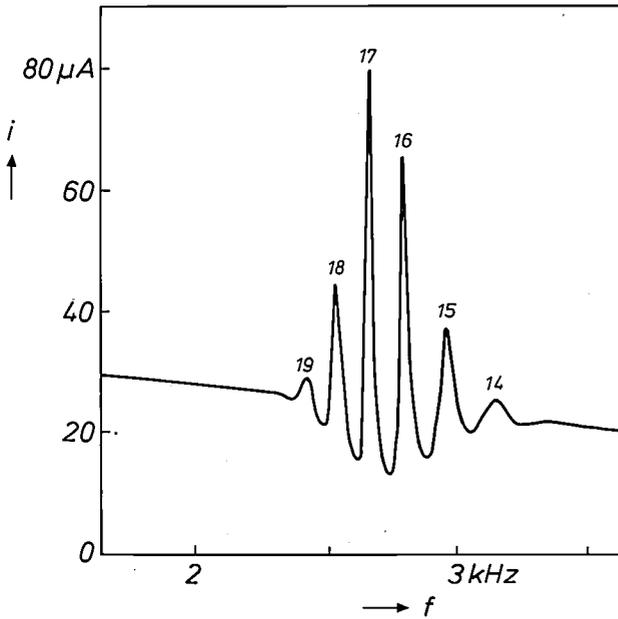


Fig. 7. Frequency spectrum of the alternating current i in the resistance. The background level of about $25 \mu\text{A}$ is the normal noise level, which is also found in the unstriated discharge. The numbers above the peaks show the length of the positive column in wavelengths at this frequency.

the field $E(x,t)$ spatially integrated from cathode to anode:

$$V(t) = \int_0^l E(x,t) dx \quad (1)$$

(with the cathode at $x = 0$ and the anode at $x = l$). It is an alternating voltage V with the same frequency as the travelling wave. The alternating current produced by this voltage in the circuit must also travel through the discharge. The current cannot be transferred directly by the travelling wave itself. The displacement of the space charge that produces the accompanying field $E(x,t)$ results in a current of no more than $1/1000$ of the observed current at most. For this reason the alternating current, like the direct current, must be transferred without spatial variations by electrons in the plasma: the column must therefore contain a purely time-dependent oscillation as well as the travelling wave. (As we shall see, this also provides the explanation for the sinusoidal spatial modulation of the wave amplitude shown in fig. 5.) The system thus contains a travelling wave that owes its continued existence to a self-maintaining oscillation. Explaining how the striations in the plasma are linked with the current in the circuit thus amounts to explaining the coupling between the travelling wave and the oscillation in the plasma. In the following section it will be shown that a description of the phenomena in

the plasma that takes both the travelling wave and the oscillation into account can provide a satisfactory explanation, leading to results that agree with the observations discussed, including the laser-type resonances.

Mathematical description

The local variations ($u(x,t)$) in the radiation from the discharge that are due to striations can be described in terms of a travelling wave:

$$u(x,t) = e^{ax} \cos(\omega t - kx), \quad (2)$$

where x is the distance from the cathode, a the rate of increase of the amplitude, ω the angular frequency and k the wave number of the travelling wave. The sinusoidal modulation of the amplitude is not taken into account. It is also assumed that the feedback is too weak for saturation of the amplitude at the anode.

The oscillation also affects the intensity. The effect of the travelling wave and the oscillation on the intensity of the radiation field can be described by the following equation:

$$u(x,t) = e^{ax} \cos(\omega t - kx) + u_r \cos(\omega t - \phi), \quad (3)$$

where u_r is the relative amplitude of the oscillation and ϕ the phase difference between the travelling wave and the oscillation at the position $x = 0$. It is easy to see that the amplitude A of the resulting wave is given by

$$A(x) = \sqrt{e^{2ax} + 2u_r e^{ax} \cos(kx - \phi) + u_r^2}. \quad (4)$$

This expression reproduces exactly what we have seen in fig. 5: a sinusoidal spatial modulation of an exponentially increasing signal. This gives an indication that the oscillation does indeed occur not only in the circuit but also in the discharge.

In fig. 5 we found a particular point where the change in intensity is always zero, irrespective of the value of the resistance included in the circuit. At position $x = x_0$ destructive interference occurs. This implies that the travelling wave and the oscillation at $x = x_0$ have the same intensity and also that their phase difference is exactly 180° . This point fits into this description as a boundary condition: $kx_0 = \pi + \phi$, and $u_r = e^{ax_0}$, which determines the coupling between the oscillation and the travelling wave.

The frequency spectrum

The resonances that occur in the current in the circuit and in the intensity of the photodiode signal can be explained in terms of the description of the wave phenomena in the discharge. As noted earlier, both

the optical intensity in the discharge and also the electric field accompanying the travelling wave satisfy an equation like equation (2). We were able to treat the plasma as an alternating voltage source supplying a voltage $V(t)$ (see page 94). Equation (1) shows that the alternating voltage between the cathode and anode is the integral of this field between $x = 0$ and $x = l$. This alternating voltage and the current it produces in the circuit must also satisfy Ohm's law:

$$V(t) = I(t)(Z + R), \quad (5)$$

where Z is the impedance^[4] of the plasma. Between the source voltage $V(t)$ (associated with the travelling wave) and the alternating current $I(t)$ in the circuit (associated with the oscillation) there must therefore be a fixed phase difference, because of Ohm's law. On integrating equation (1) it can be seen that this phase difference is directly connected with the terms in which $\sin kl$ and $\cos kl$ occur. Since the length of the tube is fixed, this phase difference can only agree with Ohm's law for particular values of k . If k_0 is a possible value at which Ohm's law is satisfied, then all other possible values are related to it by $kl = k_0l + m2\pi$, where m is an integer. As the relation between the frequency and the wave number of the travelling wave is set by the dispersion relation (a characteristic property of the plasma), it follows that the frequency of the wave and hence the frequency of the oscillation can only assume specific values. This explains the resonances shown in figs 6 and 7 in the intensity of the radiation and in the current in the circuit.

^[4] With Pekarek's model as the starting point an expression has been derived for the impedance Z in: F. C. van den Heuvel and Q. H. F. Vrethen, Striations of the convective type and feedback in low-pressure mercury/noble-gas discharges, *Phys. Fluids* 28, 3034-3039, 1985.

The argument given above demonstrates that an oscillation occurs in the discharge and in the circuit. It is now clear that a study of the discharge alone is not sufficient for a proper understanding of the striation effects in the discharge. It has been shown that Pekarek's model, which only describes the processes in the discharge, can be extended to take the effect of the electrical circuit into account. Results obtained with this model are in quantitative agreement with observations. In a convective plasma striations can be suppressed by including in the circuit a filter that gives a high attenuation at frequencies within the 'gain bandwidth' for striations. The more fundamental origin of the coupling between the travelling wave in the discharge and the current it produces in the circuit requires further investigation, however. Our measurements demonstrate that a better understanding of this coupling can be obtained by studying the point at the start of the positive column where the total intensity variation is always zero.

Summary. In a gas discharge optical instabilities, called striations, can occur in certain circumstances. In a convectively unstable plasma a study has been made of the relation between these instabilities and the electrical circuit including the discharge. Feedback via the circuit produces an alternating current that has the effect of maintaining the striations in the plasma. A better understanding of the striations can be obtained by treating the plasma and the electrical circuit as a single entity. Measurements reveal an interesting boundary condition, requiring further investigation, for the interaction between the effects in the plasma and the alternating current in the circuit. The origin of this boundary condition must lie in the discharge itself.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale, P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

A. J. den Boef	E	Interferometric laser rangefinder using a frequency modulated diode laser	Appl. Opt. 26	4545-4550	1987
D. J. Gravesteijn, H. M. van Tongeren*, M. Sens*, T. Bertens* (*Philips & Dupont Opt. Company, Eindhoven) & C. J. van der Poel	E	Phase-change optical data storage in GaSb	<i>ibid.</i>	4772-4776	1987
J. C. M. Henning & J. P. M. Ansems	E	Determination of the photoionization threshold of the deep donor in $Al_{0.33}Ga_{0.67}As:Si$	Appl. Phys. A 44	245-247	1987
D. J. Olego, J. Petruzzello, S. K. Ghandi*, N. R. Taskar* & I. B. Bhat* (*Rensselaer Polytech. Inst., Troy, NY)	N	Elastic strains in CdTe-GaAs heterostructures grown by metalorganic chemical vapor deposition	Appl. Phys. Lett. 51	127-129	1987
C. van Berkel & M. J. Powell	R	Resolution of amorphous silicon thin-film transistor instability mechanisms using ambipolar transistors	<i>ibid.</i>	1094-1096	1987
M. J. Powell, C. van Berkel, I. D. French & D. H. Nicholls	R	Bias dependence of instability mechanisms in amorphous silicon thin-film transistors	<i>ibid.</i>	1242-1244	1987
D. J. Olego	N	Effects of ZnSe epitaxial growth on the surface properties of GaAs	<i>ibid.</i>	1422-1424	1987
M. Nikoonahad	N	Differential amplitude contrast in acoustic microscopy	<i>ibid.</i>	1687-1689	1987
J. Khurgin	N	Second-order susceptibility of asymmetric coupled quantum well structures	<i>ibid.</i>	2100-2102	1987
P. Boher, M. Renaud, J. M. Lopez-Villegas, J. Schneider & J. P. Chane	L	InGaAs/Si ₃ N ₄ interface obtained in ultrahigh vacuum multipolar plasma: <i>in-situ</i> control by ellipsometry and electrical characterization	Appl. Surf. Sci. 30	100-107	1987
A. H. van Ommen	E	Diffusion of group III and V elements in SiO ₂	<i>ibid.</i>	244-264	1987
A. H. van Ommen & M. P. A. Viegars	E	Buried oxide formation in Si high-dose implantation of oxygen	<i>ibid.</i>	383-389	1987
F. F. Leopold* & F. L. van Nes* (*Inst. Perception Res., Eindhoven)		Control of data processing systems by voice commands	Behav. & Inf. Technol. 6	323-326	1987
M. Wolny, T. Aguila, P. Deconinck, D. Moroni & J. P. André	L	High-performance WN-gate MISFETs fabricated from MOVPE wafers	Electron. Lett. 23	1127-1128	1987
C. J. Mahon & G. D. Khoe	E	Endless polarisation state matching control experiment using two controllers of finite control range	<i>ibid.</i>	1234-1235	1987

- | | | | | |
|--|---|---|-----------|------|
| F. Roozeboom, A. Sikkema & L. W. Molenkamp
E | Multiple quantum well structures and high-power lasers of GaAs-AlGaAs grown by metalorganic vapor phase epitaxy (MOVPE) | Fiber & Integrated Opt. 6 | 331-345 | 1987 |
| P. R. Boudewijn & K. T. F. Janssen
E | Application of SIMS in semiconductor research | Frensenius Z. Anal. Chem. 329 | 215-219 | 1987 |
| E. K. Broadbent, M. Delfino, A. E. Morgan, D. K. Sadana & P. Maillot
S | Self-aligned silicided (PtSi and CoSi ₂) ultra-shallow p ⁺ /n junctions | IEEE EDL-8 | 318-320 | 1987 |
| S. Colak, R. Eppenga & M. F. H. Schuurmans
E,N | Band mixing effects on quantum well gain | IEEE J. QE-23 | 960-968 | 1987 |
| W. C. H. Gubbels, C. D. Hartgring, R. H. W. Salters, J. A. M. Lammerts, M. J. Tooher, P. F. P. C. Hens, J. J. J. Bastiaens, J. M. F. van Dijk & M. A. Sprokel
E | A 40-ns/100-pF low-power full-CMOS 256K (32KX8) SRAM | IEEE J. SC-22 | 741-747 | 1987 |
| E. Seevinck, F. J. List & J. Lohstroh
E | Static-noise margin analysis of MOS SRAM cells | <i>ibid.</i> | 748-754 | 1987 |
| A. J. E. M. Janssen
E | A note on 'Positive time-frequency distributions' | IEEE Trans. ASSP-35 | 701-703 | 1987 |
| P. C. W. Sommen, P. J. van Gerwen, H. J. Kotmans & A. J. E. M. Janssen
E | Convergence analysis of a frequency-domain adaptive filter with exponential power averaging and generalized window function | IEEE Trans. CAS-34 | 788-798 | 1987 |
| A. Bhattacharyya, S. N. Shabde, F. Barman & R. S. Muller (<i>Univ. of California, Berkeley, CA</i>)
S | The effects of constant-current stress on gate oxides in LDD MOSFET's | IEEE Trans. ED-34 | 942-944 | 1987 |
| S. J. Battersby & J. J. Harris
R | Millimeter-wave bulk unipolar mixer diodes | <i>ibid.</i> | 1046-1051 | 1987 |
| G. A. M. Hurkx
E | On the sidewall effects in submicrometer bipolar transistors | <i>ibid.</i> | 1939-1946 | 1987 |
| A. A. Staals*, M. C. van Houwelingen*, H. F. Huisman* (*PD Magnetics, Oosterhout) & C. A. M. Mulder
E | Localization and characterization of sub-surface particles in magnetic tape | IEEE Trans. MAG-23 | 112-114 | 1987 |
| J. J. M. Ruigrok & D. Quak (<i>Univ. of Technol., Delft</i>)
E | An alternative expression for the read flux in magnetic recording theory | <i>ibid.</i> | 1764-1766 | 1987 |
| B. J. Minnis
R | Decade bandwidth bias T's for MIC applications up to 50 GHz | IEEE Trans. MTT-35 | 597-600 | 1987 |
| W. J. A. M. Hartmann & H. M. J. Hiksipoors (<i>Philips Consumer Electron. Div., Eindhoven</i>)
E | Three-dimensional TV with cordless FLC spectacles | Inf. Disp. 3 (No. 9) | 15-17 | 1987 |
| A. F. de Jong, W. Coene* & H. Bender* (*Univ. Antwerpen)
E | HRTEM study of GaAs/AlAs interfaces: comparison of experimental, calculated and processed images | Inst. Phys. Conf. Ser. No. 87 | 9-14 | 1987 |
| A. H. van Ommen & M. P. A. Viegars
E | Structural study of the formation of a buried oxide layer by oxygen implantation | <i>ibid.</i> | 385-390 | 1987 |
| A. H. Reader, I. J. Raaijmakers & H. J. van Houtum
E | Stacking faults and precipitates in annealed and co-sputtered C49 TiSi ₂ films | <i>ibid.</i> | 523-528 | 1987 |
| C. W. T. Bulle-Lieuwma, A. H. van Ommen & J. Hornstra
E | Structural study of CoSi ₂ grown on (001) and (111) Si | <i>ibid.</i> | 541-546 | 1987 |
| W. J. Bartels
E | Computer simulation of X-ray diffraction profiles for the characterization of superlattices | <i>ibid.</i> | 599-608 | 1987 |
| S. M. Smith*, N. J. Cronin*, (*Univ. of Bath), R. J. Nicholas*, M. A. Brummell* (*Clarendon Lab., Oxford) & J. J. Harris & C. T. Foxon
R | Millimeter and submillimeter detection using Ga _{1-x} Al _x As/GaAs heterostructures | Int. J. Infrared & Millimeter Waves 8 | 793-802 | 1987 |
| Y. Genin
B | A survey of the eigenstructure properties of finite Hermitian Toeplitz matrices | Integral Equations & Operator Theory 10 | 621-639 | 1987 |

- W. M. Rabinowitz*, A. J. M. Houtsma (*Inst. for Perception Res., Eindhoven*), N. I. Durlach* & L. A. Delhorne* (**Massachusetts Inst. of Technol., Cambridge, MA*) Multidimensional tactile displays: identification of vibratory intensity, frequency, and contactor area *J. Acoust. Soc. Am.* 1243-1252 1987
82
- S. G. Nootboom (*Inst. for Perception Res., Eindhoven & Univ. Leiden*) & J. G. Kruyt (*Inst. for Dutch Lexicol., Leiden*) Accents, focus distribution, and the perceived distribution of given and new information: an experiment *ibid.* 1512-1524 1987
- J. H. M. van der Linden, P. E. Wierenga & E. P. Honig *E* Viscoelastic behaviour of polymer layers with inclusions *J. Appl. Phys.* 62 1613-1615 1987
- H. A. van Sprang & H. G. Koopman *E* Temperature compensation of the threshold voltage of a 270° twist (SBE) cell *ibid.* 1734-1738 1987
- M. Delfino, A. E. Morgan, E. K. Broadbent, P. Maillot & D. K. Sadana *S* Effect of post-silicidation annealing on TiSi₂/p⁺-n Si junctions *ibid.* 1882-1886 1987
- P. A. Breddels, H. A. van Sprang & J. Bruinink *E* Influence of dispersion on the transmission characteristics of supertwisted nematic effects in liquid-crystal displays *ibid.* 1964-1967 1987
- S. Colak, J. Khurgin, W. Seemungal & A. Hebling *N* Threshold in electron-beam end-pumped II-VI lasers *ibid.* 2633-2639 1987
- D. K. Sadana, A. E. Morgan, M. H. Norcott & Naik *S* Annealing and oxidation behavior of low-pressure chemical vapor deposited tungsten silicide layers on polycrystalline silicon gates *ibid.* 2830-2835 1987
- D. A. Cammack, R. J. Dalby, H. J. Cornelissen & J. Khurgin *N* Electron beam pumped lasing in ZnSe/ZnSSe superlattice structures grown by molecular-beam epitaxy *ibid.* 3071-3074 1987
- R. Pandya & B. A. Khan *N* Transient photocurrent response in polycrystalline silicon thin films *ibid.* 3244-3248 1987
- P. H. Oosting, J. Petruzzello & T. F. McGee *E,N* Formation of buried nitride silicon-on-insulator structures studied by Auger electron spectroscopy and transmission electron microscopy *ibid.* 4118-4123 1987
- H. Mani*, A. Joulie*, J. Bhan* (**EM-CNRS, Montpellier*), C. Schiller & J. Primot (*CNET, Bagneux*) *L* The influence of supercooling on the liquid phase epitaxial growth of InAs_{1-x}Sb_x on (100) GaSb substrates *J. Electron. Mater.* 16 289-294 1987
- F. R. de Boer*, Huang Ying-Kai* (**Univ. Amsterdam*), D. B. de Mooij & K. H. J. Buschow *E* Magnetic properties of a series of novel ternary intermetallics (RFe₁₀V₂) *J. Less-Common Met.* 135 199-204 1987
- D. B. de Mooij & K. H. J. Buschow *E* Some novel ternary ThMn₁₂-type compounds *J. Less-Common Met.* 136 207-215 1988
- R. Grössinger*, H. Kirchmayr* (**Univ. of Technol., Vienna*) & K. H. J. Buschow *E* Magnetic anisotropy in the system La₂Fe_{14-x}Co_xB and its relation to the system Nd₂Fe_{14-x}Co_xB *ibid.* 367-373 1988
- P. Schobinger-Papamantellos (*Inst. für Kristallogr. und Petrogr., Zürich*) & K. H. J. Buschow *E* Magnetic structure changes observed by neutron diffraction in the system TbGe_{1-x}Si_x (0.4 ≤ x ≤ 1.0) *J. Magn. & Magn. Mater.* 71 134-146 1988
- C. A. M. Mulder & A. A. J. M. Damen *E* The origin of the 'defect' 490 cm⁻¹ Raman peak in silica gel *J. Non-Cryst. Solids* 93 387-394 1987
- G. A. C. M. Spierings *E* Optical absorption of Ag⁺ ions in 11(Na, Ag)₂O·11B₂O₃·78SiO₂ glass *J. Non-Cryst. Solids* 94 407-411 1987
- C. A. M. Mulder *E* Defect structures in silica glass *J. Non-Cryst. Solids* 95 & 96 303-310 1987
- J. W. C. de Vries *E* Surface scattering in thin evaporated gold double layers studied by *in situ* resistivity measurements *J. Phys. F* 17 2403-2409 1987
- P. Schobinger-Papamantellos (*Inst. für Kristallogr. und Petrogr., Zürich*) & K. H. J. Buschow *E* The magnetic phase diagram of HoGe_{1-x}Si_x studied by neutron diffraction and magnetic measurements *J. Solid State Chem.* 70 249-261 1987
- P. Friedel, J.-P. Landesman, P. Boher & J. Schneider *L* Cleaning and nitridation of GaAs surfaces in multipolar plasmas investigated by *in situ* photoemission and spectroscopic ellipsometry *J. Vac. Sci. & Technol. B* 5 1129-1134 1987

- | | | | | | |
|---|---|---|---|-----------|------|
| J. van Zwol, J. van Laar, A. W. Kolfshoten & J. Dieleman | E | Effects of Ar ⁺ angle of incidence on the etching of Si with Cl ₂ and low-energy Ar ⁺ ions | J. Vac. Sci. & Technol. B 5 | 1410-1414 | 1987 |
| W. C. M. Claassen & J. Dieleman | E | An <i>in situ</i> infrared study on the interaction of oxygen plasmas with Si and fluorine plasmas with SiO ₂ surfaces | <i>ibid.</i> | 1450-1452 | 1987 |
| J. K. Annot & R. A. H. van Twist | E | A novel deadlock free and starvation free packet switching communication processor | Lecture Notes in Computer Science, Vol. 258, G. Goos & J. Hartmanis (eds), Springer, Berlin | 68-85 | 1987 |
| E. A. M. Odijk | E | The DOOM system and its applications: a survey of Esprit 415 subproject A, Philips Research Laboratories | <i>ibid.</i> | 461-479 | 1987 |
| L. Augusteijn | E | Garbage collection in a distributed environment | Lecture Notes in Computer Science, Vol. 259, G. Goos & J. Hartmanis (eds), Springer, Berlin | 75-93 | 1987 |
| A. H. Reader, A. H. van Ommen & H. J. W. van Houtum | E | Resistivity differences in C49 TiSi ₂ films formed by rapid thermal processing | Mater. Res. Soc. Symp. Proc. 92 | 177-182 | 1987 |
| A. H. van Ommen, H. J. Ligthart, J. Politiek & M. P. A. Vieggers | E | High-quality SOI by oxygen implantation into silicon | Mater. Res. Soc. Symp. Proc. 93 | 119-123 | 1987 |
| E. Gerritsen, H. J. Ligthart & T. E. G. Daenen | E | Cross-sectional TEM and corrosion studies of Al and N implanted copper | <i>ibid.</i> | 329-334 | 1987 |
| W. J. A. Goossens | E | The smectic A and the smectic C phase: a coherent molecular picture | Mol. Cryst. & Liq. Cryst. 150b | 419-445 | 1987 |
| D. J. Verschuur*, D. de Vries* (*Univ. of Technol., Delft), A. J. M. Kaizer & W. F. Drujvesteijn | E | Toepassing van de Wigner distributie bij de analyse van luidsprekerresponsies in kamers | Ned. Akoest. Genootschap nr. 87 | 31-45 | 1987 |
| R. A. Haring (IBM T. J. Watson Res. Center, Yorktown Heights, NY), H. E. Roosendaal (North-Holland Phys. Publishing, Amsterdam) & P. C. Zalm | E | On the energy and angular distribution of sputtered polyatomic molecules | Nucl. Instrum. & Methods Phys. Res. B28 | 205-213 | 1987 |
| J. P. Charlier, M. Vanbegin & P. van Dooren | B | On efficient implementations of Kogbetliantz's algorithm for computing the singular value decomposition | Numer. Math. 52 | 279-300 | 1988 |
| J. W. M. Bergmans & A. J. E. M. Janssen | E | Robust data equalization, fractional tap spacing and the Zak transform | Philips J. Res. 42 | 351-398 | 1987 |
| J. W. M. Bergmans, S. A. Rajput & F. A. M. van de Laar | E | On the use of decision feedback for simplifying the Viterbi detector | <i>ibid.</i> | 399-428 | 1987 |
| P. P. J. van Engelen & K. H. J. Buschow | E | The magneto-optical properties of Heusler alloys of the type Co _{2-x} Cu _x MnSn | <i>ibid.</i> | 429-434 | 1987 |
| P. K. Bachmann | E | Worldwide status of dispersion-modified single-mode fibres | <i>ibid.</i> | 435-450 | 1987 |
| C. Ronse | B | On rotators and shifters | <i>ibid.</i> | 451-479 | 1987 |
| R. Eppenga, M. F. H. Schuurmans & S. Colak | E | New k-p theory for GaAs/Ga _{1-x} Al _x As-type quantum wells | Phys. Rev. B. 36 | 1554-1564 | 1987 |
| D. B. M. Klaassen, C. M. G. van Leuken & K. M. H. Maessen (Univ. Utrecht) | E | 'Giant resonances' in luminescence soft-X-ray excitation spectra of phosphors | <i>ibid.</i> | 4407-4412 | 1987 |
| D. K. Maude*, J. C. Portal*, L. Dmowski* (*INSA, Toulouse & SNCI-CNRS, Grenoble), T. Foster*, L. Eaves* (*Univ. Nottingham), M. Nathan ⁰ , M. Heiblum ⁰ (IBM, T. J. Watson Res. Center, Yorktown Heights, NY), J. J. Harris & R. B. Beall | R | Investigation of the DX center in heavily doped n-type GaAs | Phys. Rev. Lett. 59 | 815-818 | 1987 |
| R. N. Bhargava, S. P. Herko & W. N. Osborne | N | Improved high-T _c superconductors | Phys. Rev. Lett. 59 | 1468-1471 | 1987 |

H. M. J. Boots	E	Inhomogeneous network formation studied by the kinetic gelation model	Physica 147A	90-98	1987
C. W. J. Beenakker	E	Two-dimensional soap froths and polycrystalline networks: why are large cells many-sided?	<i>ibid.</i>	256-267	1987
E. E. Havinga, L. W. van Horssen, W. ten Hoeve (<i>Syncom, Groningen</i>), H. Wynberg (<i>Univ. Groningen</i>) & E. W. Meijer	E	Self-doped water-soluble conducting polymers	Polym. Bull. 18	277-281	1987
M. R. Haghiri	L	Delayed decision in motion compensation technique and its application to images coding	Proc. 2nd Int. Conf. on Image processing and its applications, London 1986	205-209	1986
H. Baudry	L	Screen printing piezoelectric devices	Proc. 6th Eur. Microelectronics Conf., Bornemouth 1987	71-74	1987
P. Martin, T. Bonnet & Y. Mathieu	L	Circuit systolique pour la synthese d'images	Proc. Colloq. MARI '87, Paris - La Villette 1987	112-117	1987
C. Blanc, P. Pesqué & O. Bonnefous	L	Comparison between various beam forming techniques for ultrasound imaging	Proc. Conf. Int. Ultrasonics '87, London 1987	333-338	1987
A. Fihel & H. Sari	L	Reduced-bandwidth QAM signaling for digital radio relay systems	Proc. Eur. Conf. on Radio-relay systems, Munich 1986	173-181	1986
M. R. Haghiri & B. Boutrois	L	Tree motion estimation algorithm for video sequence	Proc. IEEE Int. Conf. on Communications, Seattle, WA, 1987	141-146	1987
M. Lamnabhi & M. Louafi	L	Burst mode operation of phase-locked loops in the presence of noise	<i>ibid.</i>	207-212	1987
S. Moridi & H. Sari	L	Effect of loop delay on the pull-in range of generalized second-order phase locked loops	<i>ibid.</i>	1041-1045	1987
H. Sari & G. Karam	L	Cancellation of power amplifier nonlinearities in digital radio receivers	<i>ibid.</i>	1809-1814	1987
P. Dawson, K. J. Moore & C. T. Foxon	R	Photoluminescence studies of type II GaAs/AlAs quantum wells grown by MBE	Proc. SPIE 792	208-213	1987
J. Braat	E	Quality of microlithographic projection lenses	Proc. SPIE 811	22-30	1987
E. W. Meijer, M. C. Raas & P. N. T. van Velzen	E	A specific 2,4-disubstituted-adamantane interaction with plasma-grown aluminium oxide. An inelastic-electron-tunnelling spectroscopy study	Recl. Trav. Chim. Pays-Bas 106	521-525	1987
J. J. van den Broek & A. G. Dirks	E	Metastable phases in binary alloy films	Scr. Metall. 21	1469-1474	1987
J. C. Brice, P. Capper, B. C. Easton, J. L. Page & P. A. C. Whiffin	R	Growth and characterisation of Cd _x Hg _{1-x} Te grown by LPE using a novel sliding boat	Semicond. Sci. Technol. 2	710-715	1987
F. J. J. Blommaert*, H. G. M. Heynen* & J. A. J. Roufs* (<i>*Inst. Perception Res., Eindhoven</i>)		Point spread functions and detail detection	Spatial Vision 2	99-115	1987
J. B. Clegg	R	Depth profiling of shallow arsenic implants in silicon using SIMS	Surf. & Interface Anal. 10	332-337	1987
P. C. Zalm	E	Quantitative sputtering	Surf. & Interface Anal. 11	1-24	1988
J. Aarts & P. K. Larsen	E	Monolayer and bilayer growth on Ge(111) and Si(111)	Surf. Sci. 188	391-401	1987
P. J. van Otterloo	E	A contour-oriented approach to digital shape analysis	Thesis, Delft	1-369	1988
P. J. M. van Laarhoven	E	Theoretical and computational aspects of simulated annealing	Thesis, Rotterdam	1-173	1988
A. G. Tangena & P. J. M. Wijnhoven	E	The correlation between mechanical stresses and wear in a layered system	Wear 121	27-35	1988



J. J. Kelly, J. E. A. M. van den Meerakker, P. H. L. Notten and R. P. Tijburg, Wet-chemical etching of III-V semiconductors,

Philips Tech. Rev. 44, No. 3, 61-74, July 1988.

III-V semiconductors like GaAs can be wet-chemically etched by three mechanisms: electrochemically with an external voltage source, electrochemically using an oxidizing agent (electroless), and chemically with a reactive compound. In some cases the etching process only proceeds when the semiconductor is exposed to light. The etch rate depends on the relative reaction rate at the semiconductor surface and the mass transfer in the solution. Other important factors are the effect of the crystal planes, the orientation of a mask with respect to these planes, and the electrical contact with other materials. Wet-chemical etching of III-V semiconductors can be used on a large scale for various applications, including the detection of crystallographic defects, the fabrication of special profiles and the selective dissolution of closely related materials in multilayer structures.

J. J. Ponjeé and P. N. T. van Velzen, Chemical modification of surfaces,

Philips Tech. Rev. 44, No. 3, 81-88, July 1988.

Bifunctional organic molecules that form bonds with a surface with one functional group give new properties to the surface with the other functional group. Adhesion to a surface or release from a surface can be improved by using this method. Preconditions for such chemical surface modification are the presence of reactive sites at the surface, the absence (because of good cleaning procedures) of impurities at these sites, and the possibility of stress relaxation on expansion and shrinkage. Subjects discussed include improvement of the adhesion between a photosensitive resist based on 'novolac' and silicon with an aminosilane, improved adhesion between noble metals and glass with a mercaptosilane, improved adhesion of silver reflecting coatings on LaserVision discs with tannin, an improvement of both adhesion and release in the replication method used for making aspheric lenses, using a methacryloxysilane and a trialkylsilane. The chemical surface modifications described here for the replication method have been investigated by IETS and TOF-SIMS. The principles of these methods are discussed and some results are presented.

A. I. Kucharska, P. Blood and E. D. Fletcher, Bistability in quantum-well lasers,

Philips Tech. Rev. 44, No. 3, 76-80, July 1988.

Hysteresis has been observed in the light-current characteristics of inhomogeneously pumped GaAs-Al_xGa_{1-x}As laser structures which were grown by molecular beam epitaxy and whose active regions consisted of 2.5 nm thick multiple quantum wells (MQW). Fast switching (< 2 ns) between low and high light-output levels was demonstrated. The observed bistability was due to saturation of the optical absorption in the passive region of these structures at high injection currents. The MQW laser devices operated in the wavelength region of the excitonic absorption and their bistability was ascribed to the decrease of this absorption at higher light intensities.

F. C. van den Heuvel, Striations in a gas discharge,

Philips Tech. Rev. 44, No. 3, 89-95, July 1988.

In a gas discharge optical instabilities, called striations, can occur in certain circumstances. In a convectively unstable plasma a study has been made of the relation between these instabilities and the electrical circuit including the discharge. Feedback via the circuit produces an alternating current that has the effect of maintaining the striations in the plasma. A better understanding of the striations can be obtained by treating the plasma and the electrical circuit as a single entity. Measurements reveal an interesting boundary condition, requiring further investigation, for the interaction between the effects in the plasma and the alternating current in the circuit. The origin of this boundary condition must lie in the discharge itself.

Ned. Philips Bedrijven B.V.
PHILIPS NATUURKUNDE LA
BILHOEFKUNST - 1
P.O. BOX 1000
5000 JA EINDHOVEN
NEDERLAND

Contents

- | | |
|---|-----|
| Magnetic domains in amorphous alloys
for tape-recorder heads
H. J. de Wit and K. Jager
<i>Research on materials for tape-recorder heads has been
stimulated by the demand for video recording
at high information densities</i> | 101 |
| Polyepoxides
E. W. Meijer
<i>Phosphorescence spectroscopy and NMR spectroscopy show how
structure and properties of polymer networks are related</i> | 110 |
| Then and Now (1938-1988) | 122 |
| Theory and practice of acoustic noise control
in electrical appliances
J. Crucq
<i>A good understanding of the origin and transmission of sound
is a great aid in the suppression of noise
in washing machines, refrigerators and
other appliances</i> | 123 |
| Scientific publications | 135 |

Magnetic domains in amorphous alloys for tape-recorder heads

H. J. de Wit and K. Jager

In modern video cassette recorders the magnetic material of the record/playback head is usually a ferrite. Ferrites are not so suitable, however, for the very high information densities necessary for recording on 8-mm tape, for example, and for digital video recording. For these applications certain amorphous cobalt-containing alloys are more promising. The magnetic behaviour of these materials is largely determined by the structure of the magnetic domains they contain. Scientists and engineers at Philips Research Laboratories have been studying ways of manipulating the domain structure to produce materials with the desired properties.

Introduction

Audio and video signals are usually recorded magnetically by drawing a tape coated with a magnetizable material past a magnetic head. The signal current generates a magnetic leakage field at the gap, and this field produces a particular magnetization in the tape. In 'axial' (or 'longitudinal') recording, which is the most common and the method we shall consider in this article, the magnetization in the tape has a preferred direction parallel to the tape surface; see *fig. 1*. The same head is generally used for playback, in which an electrical signal is induced in a coil in response to the magnetic flux from the tape, so that the recorded signal is reproduced.

A limitation on the density of the information recorded on the tape is the demagnetization due to adjacent domains magnetized in the opposite sense. The demagnetizing field should be no greater than the coercive field H_c , which is the minimum magnetic field-strength that reduces the induced magnetization to zero. As the domains become closer together, the

demagnetization increases, so that the value of H_c (the coercivity) of the tape should be increased. Applications requiring very high information densities, such as video recording on 8-mm tape and digital video recording therefore require tapes with a very high coercivity (e.g. 100 kA/m). However, if such high-coercivity tapes are to be used, the magnetic material of the tape head must be of a higher quality^[1].

To reverse the direction of magnetization in a tape the head must be capable of supplying a field of strength at least equal to the coercivity of the tape. This field depends on the magnetization M of the material. Frequently, however, the magnetic polarization J is considered rather than the magnetization. This is equal to $\mu_0 M$, where μ_0 is the permeability of free space. The maximum polarization, the 'saturation polarization' J_s of the head material, determines the maximum strength of the field produced by the head. With conventional ferrite heads J_s is about 0.5 T, not enough to reverse the magnetization in tapes with a coercivity of the

Dr H. J. de Wit and Ing. K. Jager are with Philips Research Laboratories, Eindhoven.

^[1] J. P. M. Verbunt will discuss the manufacture of tape-recorder heads in a forthcoming issue of this journal.

order of 100 kA/m. Since the same head must also be suitable for reading the information on the tape, the head material must have a relative permeability μ_r that is sufficiently high at video frequencies (up to 5 MHz), e.g. $\mu_r > 1000$. Other requirements relate to the wear of the material and its magnetostriction, i.e. the change in length resulting from a change in magnetization.

To meet these requirements the use of certain cobalt-containing alloys has been proposed [2]. One suitable composition is $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$. These alloys can be prepared in the form of an amorphous ribbon ('metallic glass'), produced by quenching a melt of the required composition by spraying it on to a spinning metal wheel. Thin amorphous films can also be deposited on a substrate by sputtering. The use of amorphous rib-

bons or sputtered thin films makes it possible to minimize the eddy currents always encountered in metallic materials. The alloys are soft magnetic materials, which means that they have a high permeability. They also have a high saturation polarization, and their magnetostriction can be kept very low with a material of appropriate composition. The requirement of a high permeability in the video frequency range can also be satisfied, but this does necessitate special precautions. These amount to 'engineering' the structure of the 'Weiss domains', the magnetic domains of different directions of magnetization found in ferromagnetic materials. The Weiss domains are separated by 'Bloch walls', in which the magnetization gradually changes from the direction in one domain to that in the neighbouring domain.

Soft magnetic materials, which are used for example in transformers, are in general characterized by the presence of large domains with walls that can readily be displaced by weak magnetic fields. For alternating magnetic fields with frequencies in the MHz range the wall velocity is generally too low to give a sufficiently high permeability. This can be substantially improved, however, if the alloy is made less magnetically soft by giving it a preferred direction of magnetization. One way of doing this is to anneal the material in a magnetic field. With this 'induced magnetic anisotropy', the positions of the domain walls and the size of the domains can be controlled in such a way that the magnetization takes place through simultaneous rotations of the magnetic moments in the domains, without the domain walls having to change position. These rotations are fast enough to give the material the required high permeability at high frequencies.

In the work described in this article a study was made of the way in which the induced magnetic anisotropy affects the domain structure and hence the magnetic properties, especially the permeability. Various cobalt-containing alloys were made in the form of amorphous ribbons or sputtered thin films [3]. The materials were annealed at different temperatures and in different orientations of the magnetic field. The domains were made visible with a polarizing microscope by means of the Kerr effect, i.e. using the change in the direction of polarization of polarized light on reflection from a ferromagnetic material, an effect that depends on the local direction of magnetization. The magnetic properties were determined in static magnetic fields and in alternating magnetic fields of different frequencies [4].

It was found that the direction of the magnetic field during annealing has a very marked effect on the domain structure [6]. The resultant domain structures

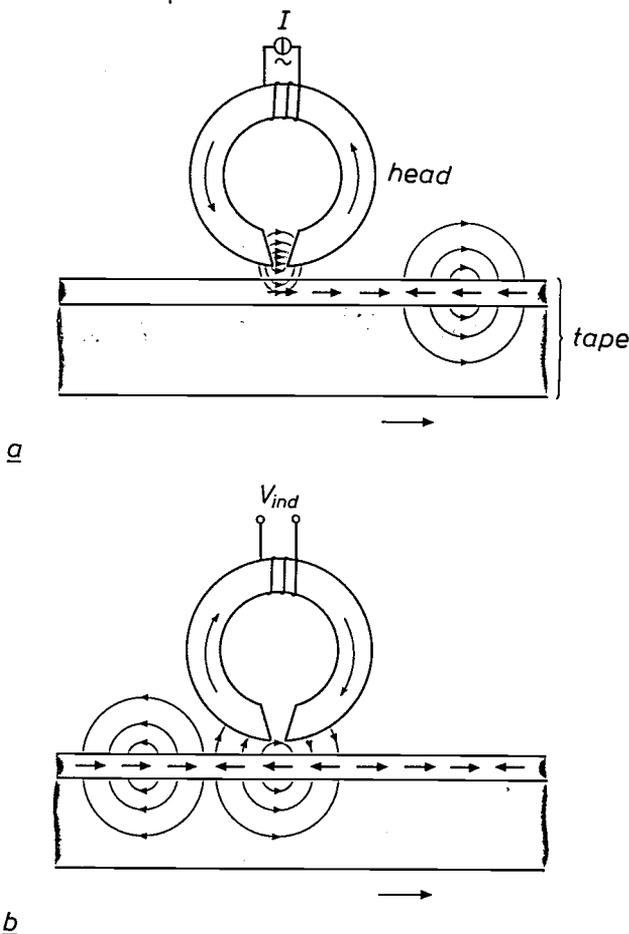


Fig. 1. *a*) Diagram representing longitudinal magnetic recording with a tape-recorder head on a moving tape coated with a magnetizable material. Magnetization directions in the materials are denoted by red arrows; magnetic lines of force outside them are denoted in blue. The signal current I generates a magnetic leakage field in the head gap via the magnetization of the head material. This leakage field records the signal on the tape by magnetizing the coating parallel to the surface of the tape. *b*) The recorded information is read with the same head. The magnetic flux from the tape magnetizes the head material, which induces a signal voltage V_{ind} in the coil in the head.

and their effect on the magnetic properties can now be explained reasonably well. In amorphous ribbons that are annealed with the magnetic field perpendicular to the axis of the ribbon the magnetization changes only through rotations of the magnetic moments. A high permeability can be obtained with such ribbons even at frequencies in the MHz range.

In this article we first describe the various domain structures that have been found in amorphous ribbons. Next we examine the magnetic behaviour of these ribbons in relation to their domain structures. Finally, we deal briefly with the changes in the magnetic properties of amorphous ribbons and sputtered thin films as a function of composition.

Control of the domain structure

Fig. 2 shows the domain structure of three amorphous ribbons 14 μm thick and about 1 cm wide, which were annealed at 300 $^{\circ}\text{C}$ in a magnetic field H_a in three different directions. The photomicrographs were made at $H = 0$, after saturating the ribbons with a magnetic field H in the longitudinal direction, i.e. parallel to the axis of the ribbon. In the ribbon annealed in a longitudinal field H_a , all the domains have '180 $^{\circ}$ walls': the magnetic moments differ by 180 $^{\circ}$ on opposite sides of the wall. The walls are parallel to H_a ; the distance between them is about 0.4 mm. The ribbon that was annealed in a transverse field H_a , i.e. with the field perpendicular to the axis and parallel to the plane of the ribbon, also has 180 $^{\circ}$ walls parallel to H_a . In this case the distance between the walls is about 0.2 mm. After an anneal with H_a perpendicular to the plane of the ribbon, a structure is found with a pattern in the form of a fine mesh; the domains are then no wider than about 3.6 μm . From the change in the Kerr effect when the ribbon is rotated with respect to the incident light beam it can be shown that the magnetization in these domains is perpendicular to the domain walls.

In the latter case a remarkable effect is observed when the ribbon is subjected to a magnetic field H whose orientation is opposite to that of the field at the original saturation. The magnetization inside the do-

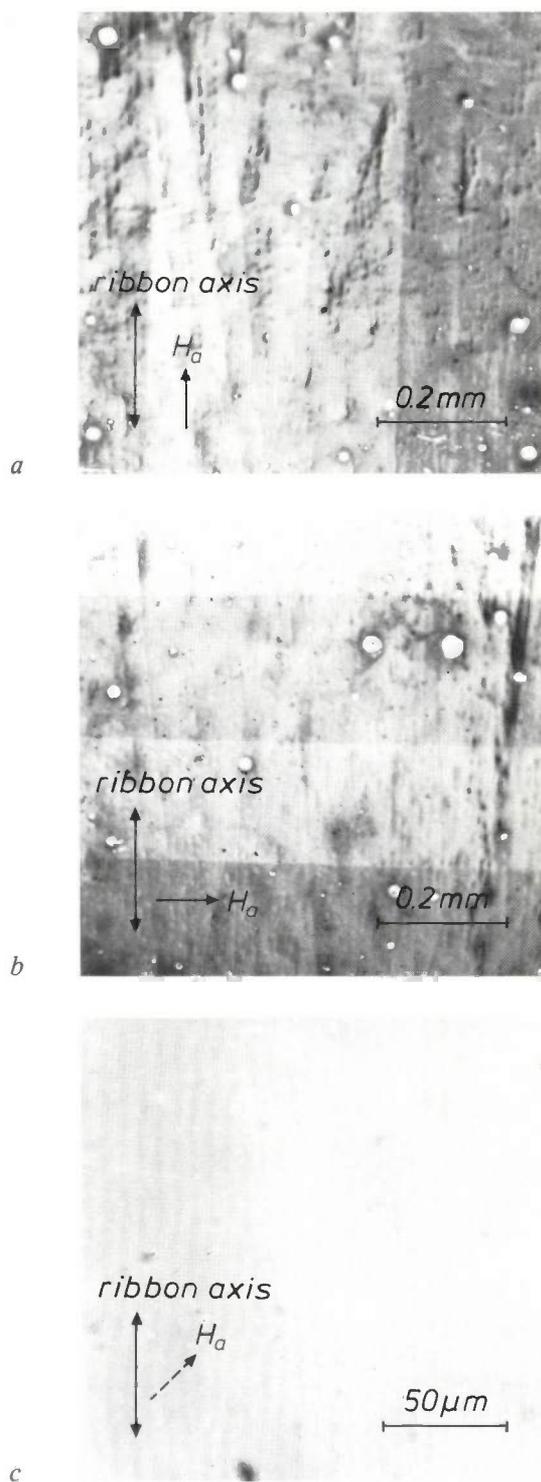


Fig. 2. Micrographs made with a polarizing microscope of domain structures in an amorphous ribbon (14 μm thick, 1 cm wide) of $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$ after annealing at 300 $^{\circ}\text{C}$ in a magnetic field H_a for three different directions. The Kerr effect has been used to make the domains visible as light and dark bands. Other contrasts are due to surface unevenness. a) With H_a (1 kA/m) parallel to the axis of the ribbon, domains are formed whose walls are about 0.4 mm apart and parallel to H_a . b) With H_a (75 kA/m) perpendicular to the axis and parallel to the plane of the ribbon, domains are again formed with walls parallel to H_a , but now about 0.2 mm apart. c) With H_a (1200 kA/m) perpendicular to the plane of the ribbon a mesh pattern is formed; the distance between the domain walls is then about 3.6 μm .

[2] The possibility of using amorphous cobalt-containing alloys for tape-recorder heads was first reported by Japanese researchers in the early eighties.

[3] The amorphous ribbons were made by Ing. H. H. Stel, the sputtered thin films by A. W. A. Bakens, Ing. H. H. Stel and Ing. J. A. M. Tolboom.

[4] Experimental work on the amorphous ribbons was also performed by Ing. A. J. C. van der Borst and Ing. G. W. Turk. The investigation of the sputtered thin films was carried out in close cooperation with Ir F. W. A. Dirne and C. H. M. Witmer; some of the magnetic measurements were performed by Ing. P. Lasinski.

[5] H. J. de Wit and M. Brouha, Domain patterns and high-frequency magnetic properties of amorphous metal ribbons, J. Appl. Phys. 57, 3560-3562, 1985.

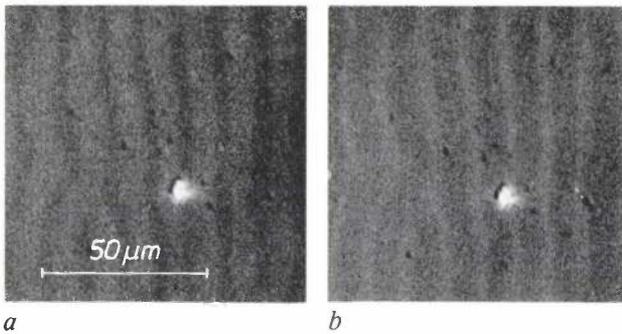


Fig. 3. Micrographs made with a polarizing microscope of an amorphous ribbon of $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$ 29 μm thick and annealed in a magnetic field perpendicular to the plane of the ribbon at 200 °C. *a*) After magnetization parallel to the ribbon axis. *b*) After applying a magnetic field in the opposite direction and stronger than the coercive field. The resulting structures are complementary.

mains then reverses its direction as soon as the field H becomes stronger than the coercivity H_c of the material. This can be seen from a comparison of the micrographs before and after H_c is exceeded; see fig. 3. A light domain changes into a dark domain and vice versa.

The domain structure after an anneal in a magnetic field perpendicular to the plane of the ribbon can be described with the well-known model^[6] shown schematically in fig. 4. According to this model the bulk of the material contains domains with 180° walls parallel to H_a . At the surface, however, 'flux-closure domains' arise, i.e. domains that form a closed magnetic flux with the domains in the bulk of the material. This results in surface domains in which the direction of magnetization is perpendicular to the domain walls. In this case the walls are parallel to the ribbon axis.

The formation of flux-closure domains occurs because it gives a reduction in the total magnetic energy density. The magnetostatic energy resulting from the emergent magnetic flux is then eliminated, and for homogeneous magnetization with a saturation polarization J_s this amounts to a reduction of the energy density by $J_s^2/2\mu_0$. For a material with $J_s = 1$ T this gives a value of 4×10^5 J/m³. This is more than three orders of magnitude greater than the energy densities in the domain walls and in the flux-closure domains.

The width of the flux-closure domains is determined by the competition between the magnetic anisotropy energy in the flux-closure domains and the energy of the 180° walls. A high anisotropy energy in the flux-closure domains implies that a great deal of energy is necessary to rotate the magnetic moments away from their preferred direction, so that narrow domains are preferentially formed. If the energy of the 180° walls is high, however, it becomes energetically advantageous for wide flux-closure domains to form. If σ_w is the wall energy per unit area (J/m²), D the thickness of the rib-

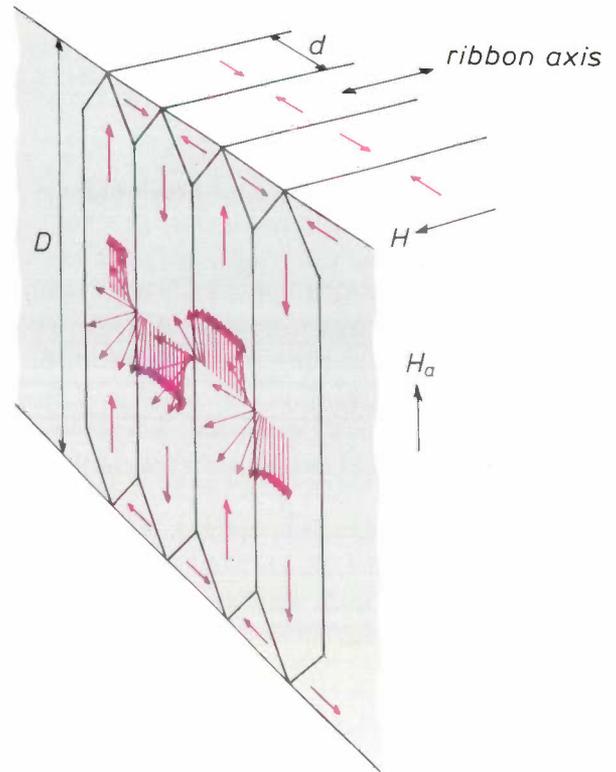


Fig. 4. Model explaining the domain structure in a ribbon of thickness D that has been annealed in a magnetic field H_a perpendicular to the plane of the ribbon and then magnetized with the magnetic field H along the ribbon axis. The red arrows indicate the direction of magnetization at $H = 0$. In the walls between neighbouring domains there is a gradual transition of the magnetization direction. At the surface there are flux-closure domains (of width d) in which the magnetization is perpendicular to the domain walls.

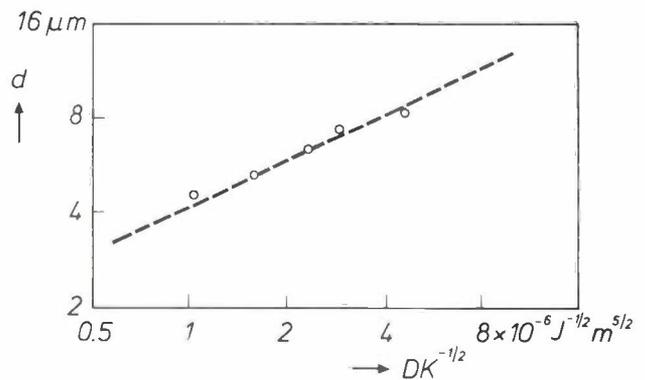


Fig. 5. Variation of the width of the flux-closure domains in an amorphous ribbon of $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$ that has been annealed in a magnetic field perpendicular to the plane of the ribbon. The width d is shown as a log-log plot against $DK^{-1/2}$, where D is the ribbon thickness and K the anisotropy constant. The measured points lie on a straight line with a slope of 0.5.

bon and K the anisotropy constant (J/m³), then the width of the flux-closure domains is given by:

$$d = (2\sigma_w D/K)^{1/2}. \quad (1)$$

The value of σ_w is determined by the anisotropy con-

stant and by the 'exchange' interaction that tries to make adjacent magnetic moments parallel:

$$\sigma_w = 4(AK)^{\frac{1}{2}}, \quad (2)$$

where A is the exchange constant (J/m). Substitution in eq. (1) then gives:

$$d = (8 A^{\frac{1}{2}} DK^{-\frac{1}{2}})^{\frac{1}{2}}. \quad (3)$$

Fig. 5 shows a log-log plot of the measured width of the flux-closure domains against $DK^{-\frac{1}{2}}$. The measured points lie on a straight line with a slope of 0.5, in agreement with the relation that follows from eq. (3):

$$\log d = \frac{1}{2} \log 8A^{\frac{1}{2}} + \frac{1}{2} \log DK^{-\frac{1}{2}}. \quad (4)$$

The value of the exchange constant that can be derived from the straight line in fig. 5, $A = 4.9 \times 10^{-12}$ J/m, is comparable with the values reported for other similar materials [7].

Our results indicate that the model with flux-closure domains gives a reasonable explanation of the observed domain structure. It can also explain the reversal of the directions of magnetization in the domain for an increasing magnetic field in the opposite direction (fig. 3). Under the influence of this field the magnetic moments in the flux-closure domains tend to align themselves towards the magnetic field. This can be seen as an attempt to form domains with 360° walls: adjacent domains acquire the same direction of magnetization, while they are separated by walls with different directions of magnetization. When the magnetic field has reached a strength corresponding to the coercivity of the material, however, this becomes energetically so unfavourable that the direction of magnetization in all the domains reverses.

Effect on the magnetic properties

The three ribbons whose domain structures are given in fig. 2 show striking differences in their magnetic behaviour. In fig. 6 the 'static' magnetization curves for these ribbons are plotted for a longitudinal magnetic field H . The ribbon annealed with a longitudinal magnetic field H_a gives a curve with a very steep central portion. This means that the permeability is very high: a weak magnetic field H is sufficient to cause

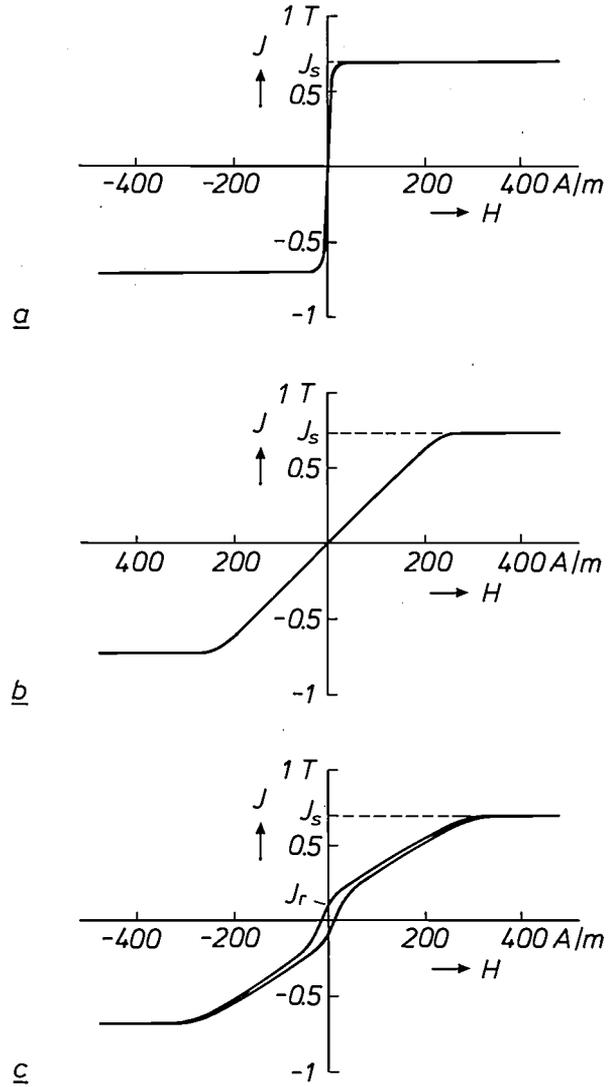


Fig. 6. 'Static' magnetization curves for the three ribbons in fig. 2, annealed in a magnetic field H_a in three different directions. The magnetic polarization J is plotted against the longitudinal magnetic field H , i.e. parallel to the ribbon axis. J_s , saturation polarization. a) The ribbon annealed in a longitudinal magnetic field gives a curve with a very steep central portion, with virtually no hysteresis. b) The ribbon annealed in a transverse magnetic field gives a curve with a less steep central portion, again with virtually no hysteresis. c) The ribbon annealed in a magnetic field perpendicular to the plane of the ribbon gives a curve with an even less steep central portion and indicating distinct hysteresis. The remanence J_r is also shown for this case.

very strong magnetization. This is typical of a magnetization process that comes about mainly through displacements of the domain walls, where the domains with the same direction of magnetization as the applied field grow at the expense of the domains with the magnetization direction in the opposite direction. The magnetization curve shows virtually no hysteresis; the displacement of a domain wall is a reversible process.

The ribbon annealed with a transverse field H_a also gives a curve with no significant hysteresis. The permeability, however, is now much lower. This indicates

[6] This model was first described by L. Landau and E. Lifshitz, On the theory of the dispersion of magnetic permeability in ferromagnetic bodies, Phys. Z. Sowjetunion 8, 153-169, 1935. Further work on the model was done by C. Kittel, Physical theory of ferromagnetic domains, Rev. Mod. Phys. 21, 541-583, 1949.

See also the review article by C. Kittel and J. K. Galt, Ferromagnetic domains, Solid State Physics, part 3, Academic Press, New York 1956, pp. 437-564.

[7] H. Kronmüller, M. Fähnle, M. Domann, H. Grimm, R. Grimm and B. Gröger, Magnetic properties of amorphous ferromagnetic alloys, J. Magn. & Magn. Mater. 13, 53-70, 1979.

that it is not the wall displacements that are important now, but the rotations of the magnetic moments towards the external field H .

The ribbon annealed with the field H_a perpendicular to the plane of the ribbon gives an even lower permeability. Here again the magnetization is determined mainly by the rotations of the magnetic moments. Unlike the other two cases, however, there is now a distinct hysteresis. This is attributable to a change in the mesh pattern. The hysteresis is related to the occurrence of remanence, i.e. the residual magnetization when the magnetic field is reduced from saturation to $H = 0$. The remanence can easily be explained from the model in fig. 4: after saturation and the return to $H = 0$, the magnetization in the domain walls still has a component in the direction of the original magnetic field.

The hysteresis found after annealing in a magnetic field perpendicular to the plane of the ribbon depends on the annealing temperature and the ribbon thickness. Some magnetization curves for different annealing temperatures and ribbon thicknesses are shown in fig. 7. When the annealing temperature is increased the central portions of the curves become less steep. The reduction in the permeability is due to the increase in the magnetic anisotropy. The remanence decreases at higher annealing temperatures and with thicker ribbons. This is in agreement with the domain-structure model. This model indicates that the ratio of the remanence J_r to the saturation polarization J_s will be approximately equal to the ratio of the width δ of the domain walls to the width d of the domains. The width of the walls is given by $\delta = \pi A^{1/2} K^{-1/2}$ [8] and the width of the domains is given by eq. (3). It follows from this that

$$J_r/J_s \approx \frac{\pi A^{1/2} K^{-1/2}}{(8D)^{1/2}} \quad (5)$$

The reduction in J_r at higher annealing temperatures can be attributed to the stronger magnetic anisotropy, i.e. to the higher value of K . The lower J_r found with a higher ribbon thickness D is also explained by this equation. Calculations for different values of K and D , with the value of A derived from fig. 5, give J_r/J_s -values that are about half those of the experimental values. A possible explanation for this is that the magnetization in the flux-closure domains also continues to have a component in the original direction of the magnetic field.

An important property for application in tape-recorder heads is the permeability in alternating magnetic fields, especially at high frequencies (> 1 MHz). The frequency-dependence of the permeability, measured at a field amplitude of 5 A/m, is shown in fig. 8 for the three ribbons of fig. 2. The ribbon annealed in a longi-

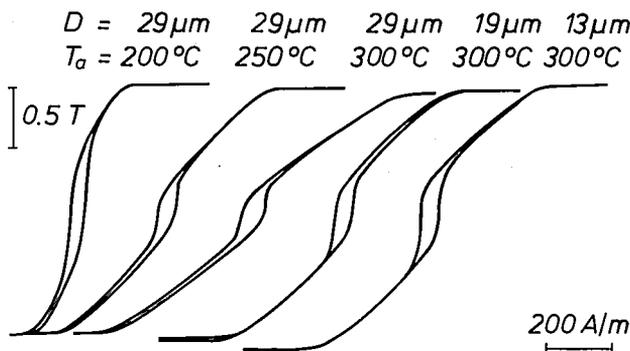


Fig. 7. Magnetization curves for an alternating magnetic field at a frequency of 1.3 Hz, measured for amorphous ribbons of thickness D annealed in a magnetic field perpendicular to the plane of the ribbon at a temperature T_a . An increase in T_a results in a curve with a less steep central portion and a lower remanence. Reducing D gives a higher remanence.

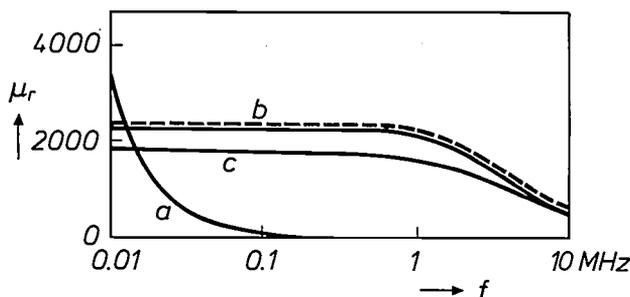


Fig. 8. Relative permeability μ_r as a function of frequency f for the three ribbons in fig. 2. The ribbon annealed in a longitudinal magnetic field (curve a) has a very high permeability at low frequencies; above 0.01 MHz, however, the permeability decreases sharply. The ribbon annealed in a transverse magnetic field (curve b) has a permeability that is constant up to about 1 MHz, then decreases. The measured curve agrees well with a calculation (dashed line) in which it was assumed that the magnetization is due to rotations of the magnetic moments in the domains, and that otherwise only eddy currents are involved. The ribbon annealed in a magnetic field perpendicular to the plane of the ribbon (curve c) has a somewhat lower constant μ_r up to about 1 MHz; again there is a decrease at higher frequencies.

tudinal magnetic field H_a gives by far the highest permeability at low frequencies (10 kHz). At higher frequencies, however, there is a very marked decrease. This is characteristic of magnetizations due to domain-wall displacements: the wall displacements are too slow to follow the magnetic field at high frequencies. In the other two cases the permeability is fairly constant up to about 1 MHz. The magnetization here is produced by rotations of the magnetic moments, which do take place fast enough to follow the magnetic field. The decrease in the permeability at frequencies in the MHz range can be attributed to eddy currents. Since the material is an electrical conductor, a rapidly varying magnetic field produces circulating currents. These produce a magnetic field in the opposite sense to the external field. As a result this field does not penetrate so far into the bulk of the material, so that the changes in magnetization are smaller.

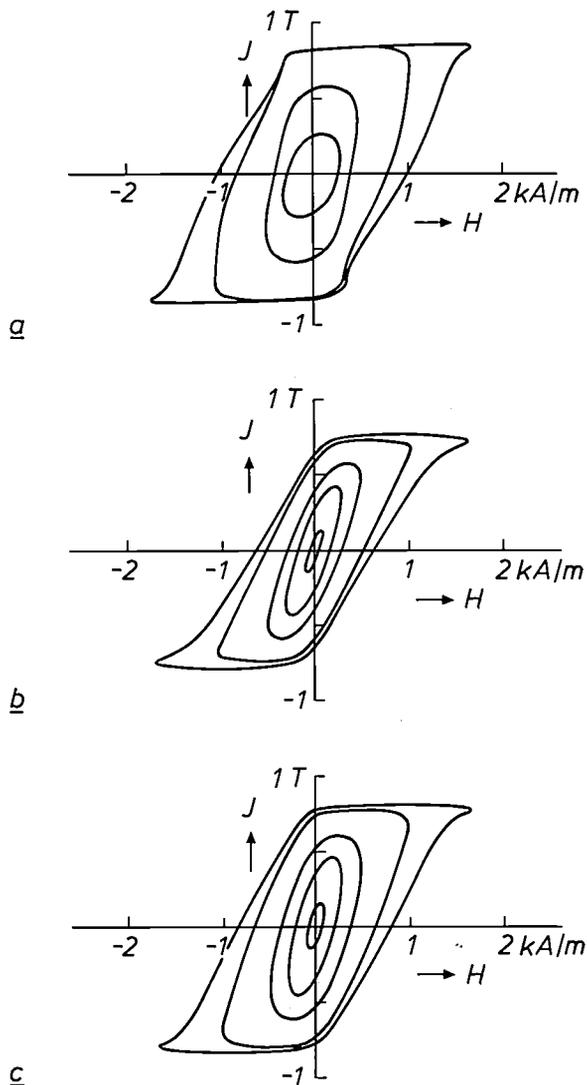


Fig. 9. Hysteresis loops for an alternating magnetic field at a frequency of 4.6 MHz for the three ribbons in fig. 2. Annealing in a longitudinal magnetic field (a) gives much broader loops than annealing in a transverse magnetic field (b) and in a magnetic field perpendicular to the plane of the ribbon (c). In (c) the loops are again broader than in (b).

The effect of eddy currents on the permeability has also been calculated^[5] for the ribbon annealed in a transverse magnetic field (curve *b* of fig. 8). It was assumed in these calculations that the magnetization is entirely due to uniform rotations, corresponding to a static magnetization curve with a constant slope and virtually no hysteresis (fig. 6*b*). It was also assumed that there are no other losses. The frequency-dependence of the permeability can then be calculated exactly^[9]. The result, which is also shown in fig. 8, is in very good agreement with the experimentally determined frequency-dependence of the permeability. This indicates that the assumptions about the rotation of the magnetic moments and the eddy currents are correct.

At high frequencies all three ribbons give hysteresis loops that are wide compared with the static magnetization curves. Fig. 9 shows a few loops measured at a frequency of 4.6 MHz. Here again the difference between wall displacements and rotations is clearly apparent. In the first case, strong fields are required to move the walls through the material, and therefore the loops in fig. 9*a* are distinctly wider than those in fig. 9*b* and *c*. The loops in fig. 9*c* are wider than those in fig. 9*b* because of the greater hysteresis in the static magnetization curve (fig. 6).

Composition-dependent changes in magnetic properties

With applications in tape-recorder heads in mind we investigated two kinds of alloy. One is formed by the amorphous ribbons discussed above, consisting of alloys of Co, Fe, Si and B. The magnetostriction of these materials can be made negligibly small by using the appropriate ratio of Co to Fe. The other kind comprises amorphous alloys of Co, Nb and Zr that are sputtered on a glass substrate. Here the magnetostriction can be suppressed via the ratio of Nb to Zr.

Besides the magnetostriction, the induced magnetic anisotropy is also an important property in the search for the most suitable composition. As we have shown, this largely determines the domain structure, which has a particularly important effect on the frequency-dependence of the permeability. The requirement for a high saturation polarization also has to be met, of course. We have investigated these properties in both of the alloys mentioned, not only varying the ratios of the different elements but also adding small amounts (about 2 at.%) of other elements. Some results of this investigation will now be briefly discussed.

Magnetostriction

We determined the magnetostriction λ_s of the ribbons by measuring the anisotropy constant K with and without a tensile stress σ . For a difference ΔK the value of λ_s is given by ^[8]:

$$\lambda_s = \frac{2}{3} \Delta K / \sigma. \quad (6)$$

Immediately after manufacture, i.e. before annealing, an amorphous ribbon with a composition $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$ gives no measurable magnetostriction: $\lambda_s \approx 0$. This changes, however, when the ribbon is first annealed in a magnetic field to induce the required magnetic anisotropy. Fig. 10 shows how the magnetostriction depends on temperature during annealing in a transverse magnetic field. After annealing at 300 °C (for example) the magnetostriction is no longer negligible: $\lambda_s \approx 3 \times 10^{-7}$. If a ribbon with

^[8] See for example S. Chikazumi, *Physics of magnetism*, Wiley, New York 1964.

^[9] See for example E. Olsen, *Applied magnetism, a study in quantities*, N.V. Philips Gloeilampenfabrieken, Eindhoven 1966, pp. 41-46.

$\lambda_s \approx 0$ is to be obtained in this case, then the ratio of Co to Fe must be changed slightly to obtain the compositions $\text{Co}_{70.8}\text{Fe}_{4.2}\text{Si}_{15}\text{B}_{10}$.

Fig. 10 also shows the effect of the annealing temperature after the addition of other elements. The ad-

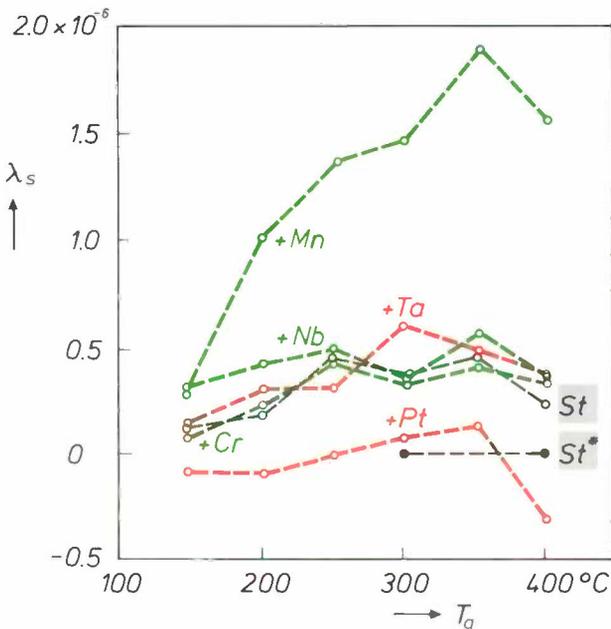


Fig. 10. The magnetostriction λ_s of some amorphous ribbons as a function of the temperature T_a at which they were annealed in a transverse magnetic field. Both for a ribbon with the 'standard composition' $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$ (St) and for ribbons that also contain 2 at. % of Pt, Ta (red), Cr, Nb or Mn (green), the magnetostriction has its maximum value at an annealing temperature of 300 to 350 °C. The addition of Mn gives a much higher λ_s , the addition of Pt gives a very much lower value. To obtain $\lambda_s \approx 0$ at $T_a = 300$ °C the ratio of Co to Fe can also be modified to give the composition $\text{Co}_{70.8}\text{Fe}_{4.2}\text{Si}_{15}\text{B}_{10}$ (St*).

dition of Mn results in a substantial increase in λ_s , particularly at high annealing temperatures. The addition of Pt, on the other hand, results in a marked decrease in λ_s , while adding other elements (Nb, Ta or Cr) has relatively little effect. In general, λ_s increases with the annealing temperature up to a value between 300 and 350 °C. At even higher temperatures a decrease occurs, probably because the Curie point of these materials is then exceeded.

The magnetostriction of the thin films sputtered on a non-magnetic substrate was determined with an experimental arrangement designed at these Laboratories by M. Brouha. In this arrangement a specimen (i.e. substrate with magnetic layer) is placed in a magnetic field that rotates at a frequency of 5 Hz. One end of the specimen is fixed and the other is left free. The rotation of the field and the magnetostriction make the magnetic layer periodically change in length; this leads to curvature of the specimen and an excursion of the free end. The amplitude of this excursion is measured with a photosensitive sensor to an accuracy of better than 0.5 nm. The smallest value of the magnetostriction that can be measured with this arrangement is about 10^{-8} .

Annealing in a magnetic field also causes a change in the magnetostriction of thin films of alloys with Co, Nb and Zr. An example is shown in fig. 11. Here the measured magnetostriction is plotted against the ratio of Nb to Zr for non-annealed films and for films annealed at 390 °C in a transverse magnetic field. This anneal causes a positive shift in the values of λ_s . In this case λ_s is reduced to about zero when the composition is changed from $\text{Co}_{86}\text{Nb}_{9.3}\text{Zr}_{4.7}$ to $\text{Co}_{86}\text{Nb}_{10.5}\text{Zr}_{3.5}$.

Induced anisotropy and saturation polarization

The magnetic anisotropy induced in the amorphous ribbons that we investigated is generally significantly less than that in the sputtered films; see fig. 12. In both kinds of material the magnetic anisotropy constant K after annealing in a transverse magnetic field was found to be highest when the annealing temperature was in the region of 300 °C. The addition of another element (e.g. Ta) usually causes a slight decrease in K , but a 'magnetically active' element like Pt gives a marked increase. The explanation of these effects will not be considered here.

The effect of the composition on the saturation polarization J_s and the anisotropy constant K is shown in fig. 13 for a large number of ribbons and sputtered films that were annealed at 300 °C in a transverse magnetic field. A higher value of J_s is generally associated with a higher value of K . The relation between J_s and K is important because these are the quantities that to a first approximation determine the

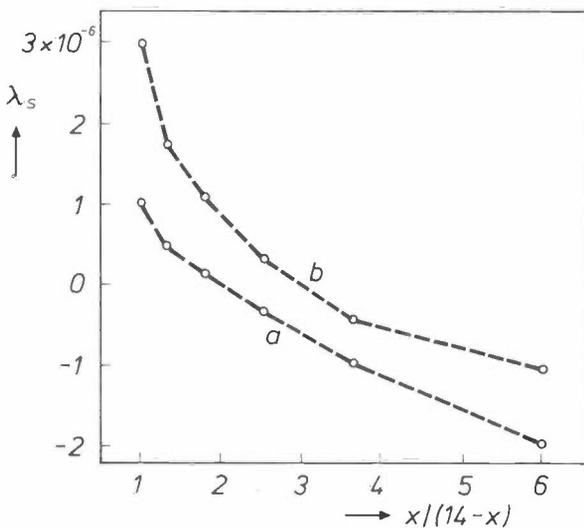


Fig. 11. The magnetostriction λ_s of sputtered thin films of the standard composition $\text{Co}_{86}\text{Nb}_x\text{Zr}_{14-x}$ as a function of the ratio of Nb to Zr, without annealing (curve a) and after annealing in a transverse magnetic field at 390 °C (curve b). In both cases an increase in this ratio gives a decrease in λ_s . Annealing causes a positive shift in λ_s .

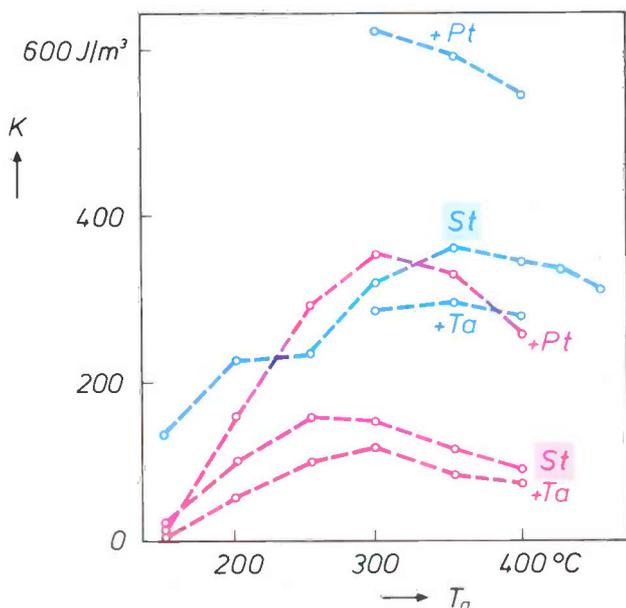


Fig. 12. The magnetic anisotropy constant K of amorphous ribbons (red) and sputtered thin films (blue), annealed in a transverse magnetic field, as a function of the annealing temperature T_a . The curves relate to the standard composition with no addition (St) and to this composition with 2 at.% of another element added. The highest K -values are reached with T_a at about 300°C . The ribbon with the standard composition $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$ has a much lower K -value at all annealing temperatures than the sputtered films with the standard composition $\text{Co}_{87.3}\text{Nb}_{8.8}\text{Zr}_{3.9}$. In both cases the addition of Ta gives a slight decrease in K , and the addition of Pt gives a marked increase.

relative permeability μ_r of the material:

$$\mu_r = \frac{J_s^2}{2\mu_0 K} \quad (7)$$

The curve that corresponds to this equation for $\mu_r = 1000$ is also shown in fig. 13. For measured points below this curve the value of μ_r is greater than 1000, whereas the points above it correspond to smaller values of μ_r . In the ribbons the addition of most elements results in a distinct decrease in J_s ; the effect on K is of minor significance, however. The addition

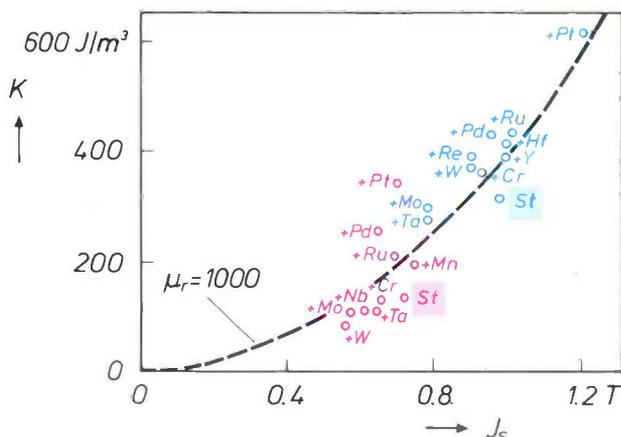


Fig. 13. The magnetic anisotropy constant K as a function of the saturation polarization J_s for amorphous ribbons (red) with the standard composition $\text{Co}_{70.3}\text{Fe}_{4.7}\text{Si}_{15}\text{B}_{10}$ and with the addition of 2 at.% of another element, and for sputtered thin films (blue) with the standard composition $\text{Co}_{87.3}\text{Nb}_{8.8}\text{Zr}_{3.9}$ and with the addition of 2 at.% of another element. The materials were annealed at 300°C in a transverse magnetic field. The dashed line corresponds to eq. (7), with $\mu_r = 1000$. Most measured points lie on or near this line. The sputtered films generally show not only a higher anisotropy but also a substantially higher saturation polarization than the ribbons. The additive that gives the strongest effect is Pt.

of Pt, on the other hand, has a much stronger effect on K than on J_s . The sputtered films have a higher saturation polarization than the ribbons, and in most cases their anisotropy is also larger. The highest values of J_s and K were obtained with films containing Pt.

Summary. Certain amorphous cobalt-containing alloys with high saturation polarizations (about 1 T) show promise for applications in tape-recorder heads for video recording at high information densities. The relative permeability μ_r , which has to be very high ($\mu_r > 1000$) for reproducing the stored information, is largely determined by the structure of the magnetic domains. This can be controlled by annealing the alloys in a magnetic field, thus inducing a magnetic anisotropy. The effect on the permeability and the other magnetic properties depends on the direction of the magnetic field and the temperature during annealing, on the form of the material — spun ribbon or sputtered thin film — and on the precise composition.

Ned. Philips Bedrijven B.V.
 PHILIPS HATUUTREKENDIG LAB.
 BIBLIOTHEEK 107-1
 P.O. BOX 1000
 5000 JA Eindhoven
 NEDERLAND

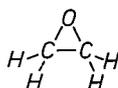
Polyepoxides; formation and properties of their network structure

E. W. Meijer

Polyepoxides are polymer materials widely used in the electronics industry, and there is much interest today in a better understanding of their structure at the molecular level. The article below discusses two methods — phosphorescence spectroscopy and nuclear magnetic resonance spectroscopy — that can be used to give 'a closer view' of polyepoxides.

Linear-chain polymerization

Polyepoxides are obtained by the polymerization of epoxide monomers [1]. Epoxides have the structural feature of a three-membered ring, consisting of two carbon atoms and one oxygen atom. The simplest monomer is ethylene oxide:



The ideal bond angles of carbon and oxygen cannot be realized in a three-membered ring. The bonds in an epoxide ring are forced to take up 'unusual' angles averaging 60° (instead of about 109°), which results in a structure possessing an internal ring strain and a corresponding strain energy of about 60 kJ/mol.

A chemical reaction that opens the epoxide ring (*fig. 1*) gives a gain in energy because of the loss of ring strain. This energy gain provides the thermodynamic basis for the 'ring-opening polymerization' of epoxides [2].

Apart from the fulfilment of this thermodynamic condition, a good kinetic route has to be found along which the polymerization can take place sufficiently rapidly. This usually requires a catalyst or initiator. By a catalyst we mean a compound that greatly increases the rate of polymerization but is not used up in the reaction. An initiator may be necessary to start the polymerization in some cases, but does take part in the reaction, usually becoming incorporated in the polymer.

Various monomers are of importance in the applications of epoxide-based polymers. In these complex monomers one or more hydrogen atoms in ethylene oxide are replaced by other atoms or groups of atoms. Steric interaction between these substituents in the epoxide ring increases the ring strain. However, this may not give an increased reactivity in polymerizations; this depends on the nature of the reaction mechanisms for the various types of polymerization.

The polymerization can take place in the form of an addition reaction between compounds of the same kind, 'chain polymerization', or in the form of a step reaction — a succession of individually catalysed reactions between different kinds of compounds — 'condensation polymerization' [2].

Chain polymerizations can be initiated by activating the epoxide ring of a monomer by a catalyst in such a way that it can react with a non-activated monomer. The resultant dimer remains activated, and so do the trimer, tetramer and so on. In this way hundreds or thousands of monomers can be concatenated to form a polymer in the 'propagation step' of the polymerization. In a 'termination reaction' the chain is deactivated and the catalyst is re-formed. It can then be used again to catalyse the formation of another chain. The catalyst can be an acid (a cationic catalyst) or a base (an anionic catalyst). *Fig. 2a* shows the progress of an acid-catalysed polymerization. The polymerization of epoxides started by an initiator can be regarded as a repetitive reaction in which the initiator is lengthened by one monomer unit at each repetition. The initiator may be for example an organo-

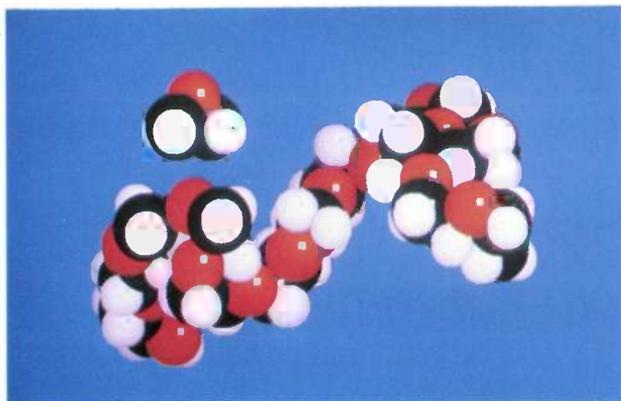


Fig. 1. Three-dimensional computer model of ethylene oxide (a) and polyethylene oxide (b). The spheres represent the Van der Waals radii of the various atoms. The models were produced by 'molecular-mechanics optimization'.

metallic compound (coordinative initiation). In this case the polymerization is terminated without the initiator being re-formed.

The reaction rates of these chain polymerizations may differ considerably. For instance, polymerizations catalysed by an acid like BF_3 may often have to be carried out at a temperature below -50°C to sustain the reaction, which is exothermic. On the other hand, polymerizations catalysed by a base, such as imidazole, only reach an acceptable rate at a temperature of about 100°C .

The second form of polymerization, condensation polymerization, is of great practical importance, since it is the basis for the widely used epoxy resins. Although it is mainly polymer networks that are produced in this way, the principle will be discussed here for linear chains. To synthesize such polymers a mixture of two kinds of compounds is used. One kind has two epoxide rings and the other contains two functional groups that can react with epoxide rings, e.g. amines, hydroxides or phenolic hydroxyl groups. Multiple repetition of the ring opening, catalysed each time, produces a polymer in a succession of steps (fig. 2b).

The macroscopic (physical, chemical and mechanical) properties of the polymeric end-product depend on the way in which the polymers are formed. In chain polymerization a cyclic ether (an epoxide) is converted into a linear — and much more stable — ether, built up by a regular repetition of a C-C-O unit in the main chain of the polymer (fig. 2a). Such regularly structured polymers have a strong tendency to crystallize. In condensation polymerizations of epoxides the main chain always contains polar side chains in the form of hydroxyl groups. These polymers are not so highly regular, and therefore tend to have the properties of amorphous glassy systems.

For completeness it should be mentioned that there are also combined forms of the two modes of polymerization described above, which can result in highly complicated systems.

Formation of three-dimensional networks

A recent article in this journal dealt with the applications of polymer chemistry in the electrical industry^[3]. An important feature in this chemistry is making polymer material insoluble. This can be done by linking linear polymers together, or by simultaneously polymerizing and cross-linking multifunctional monomers^[4]. The latter method is widely used for making polyepoxides from bisepoxides containing two epoxide rings. Some examples are shown in fig. 3.

The many and various ways in which the chemical structure of these bisepoxides can be modified, and the diversity in polymerization reactions, ensure that materials can be produced to fit the particular specifi-

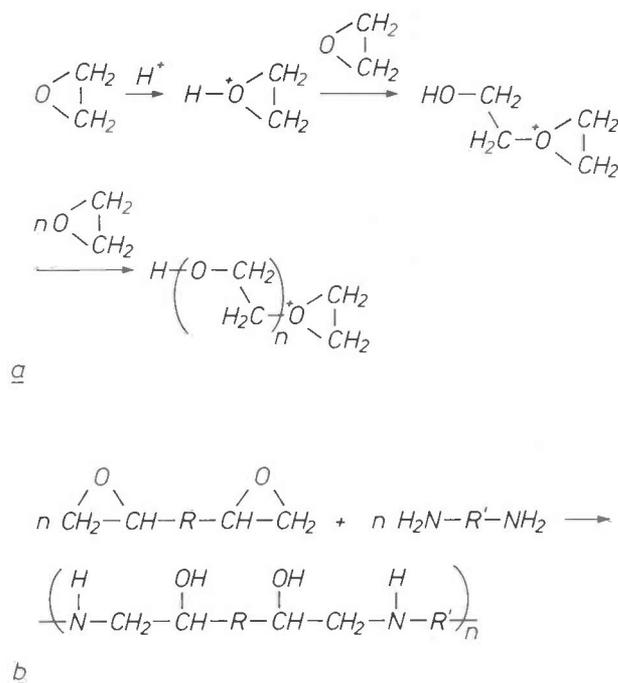


Fig. 2. a) Mechanism of an acid-catalysed chain polymerization of epoxides. b) Reactions for a condensation polymerization of a bisepoxide with a diamine. R and R' represent linking chains.

- [1] Y. Ishii and S. Sakai, in: Ring-Opening Polymerization, K. C. Frisch and S. L. Reegen (eds), Marcel Dekker, New York 1969, pp. 13-109.
- [2] S. Inoue and T. Aida, Cyclic ethers, in: Ring-Opening Polymerization, part 1, K. J. Ivin and T. Saegusa (eds), Elsevier, Barking 1984, pp. 185-298.
- [3] L. K. H. van Beek, Polymer chemistry in the electrical industry, Philips Tech. Rev. 42, 149-159, 1986.
- [4] J. G. Kloosterboer, G. J. M. Lippits and H. C. Meinders, Photopolymerizable lacquers for LaserVision video discs, Philips Tech. Rev. 40, 298-309, 1982.

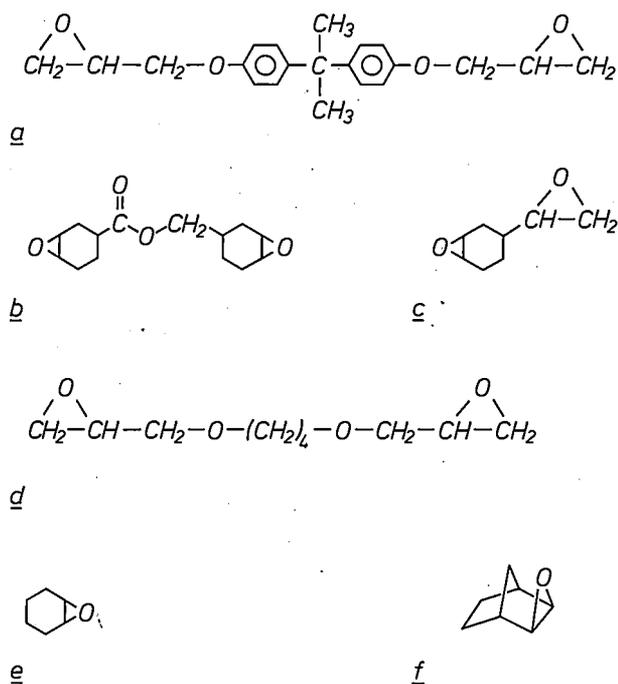


Fig. 3. Four examples of widely used bisepoxides and two monoepoxides that we used as models. *a*) bisphenol-A diglycidyl ether, *b*) (3,4-epoxycyclohexyl)-methyl 3',4'-epoxycyclohexane carboxylate, *c*) 1-epoxyethyl-3,4-epoxycyclohexane, *d*) 1,4-butanediol-diglycidyl ether, *e*) cyclohexene oxide, *f*) exo-2,3-epoxy-norbornane.

cations for any application. For most purposes condensation polymerization of two different compounds is used, because this is the method that offers the most variations in molecular structure. Examples are the epoxy resins for the encapsulations of ICs and discrete components, laminated resins for printed-circuit boards, and many adhesives, some of them conductive.

The formation of a three-dimensional network is described in this article by taking the chain polymerization of bisepoxides as an illustration. The illustration has more general validity, however. Polyepoxide networks are formed in a series of repeated reactions in which the low-viscosity bisepoxide is gradually converted into a hard end-product. This increase in viscosity during the reaction reduces the mobility of the monomers and their functional groups. The effect is intensified after the 'gel point' has been passed. The reduced mobility of the monomers and their functional groups has little influence, however, on the reaction rate of polymerization and cross-linking. Because the collisions between molecules become slower, fewer collisions are required on average to cause a reaction^[6]. The rate of conversion is not in fact limited until 'vitrification' occurs. At this point the mobility of the functional groups that have not yet

reacted is virtually zero because of the well-advanced cross-linking and the associated loss of free volume. In this 'glassy' state the reaction comes to a virtual standstill, without all the epoxides necessarily having reacted. The transition from the initial situation to the end-product is shown schematically in fig. 4.

In practice it is often advantageous if the polyepoxide networks are formed rapidly at the lowest possible temperature. Since this conflicts with other practical requirements — for durable and workable monomers without premature reaction — use is often made of 'latent catalysts'. In a latent catalysis an inert compound first has to be converted into a highly active catalyst before the polymerization can proceed. Latent catalysts can be formed by both thermal and photochemical methods.

Proper control of polymer formation requires not only a general understanding of what happens during the formation process but also a closer understanding of events on a molecular scale. Ultimately, local chemical variables determine the macroscopic physical and mechanical properties. It would be desirable to know the degree of homogeneity of the structure during the polymerization process, the precise reaction mechanisms in the different local situations, the probability of isomers being formed in the various mechanisms, and — not least — the ultimate mobility of individual polymer segments and the associated ultimate degree of conversion.

The insolubility of polymer networks makes it difficult to acquire this information, and in some cases even impossible. In the following sections two examples will be discussed that illustrate the way in which an investigation of this kind can be tackled.

Both examples deal with the formation and structure of polyepoxide networks, formed by chain polymerization with simultaneous cross-linking of bis-

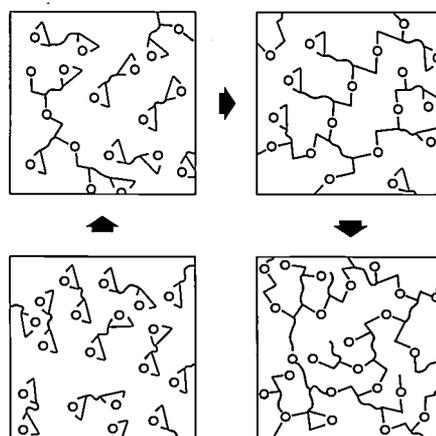


Fig. 4. Diagram of the formation of a polyepoxide network from a bisepoxide.

epoxides. In the first example the way the polymerization process takes place is broadly known and we can therefore concentrate on more detailed aspects, such as the molecular mobility. It will be shown how phosphorescence spectroscopy can be used to monitor the change in the molecular mobility of polymer segments during the formation of a network. If the mechanism of polymerization is entirely unknown, however, this must be established first. The second example will show how the mechanism of network formation for model compounds can be studied by means of nuclear-magnetic-resonance spectroscopy, a powerful method for the spectroscopic investigation of organic molecules [6].

Molecular mobility of polyepoxide networks [7]

Phosphorescence spectroscopy as an investigative method

A diaryl-iodonium compound can be used as the latent catalyst or photoinitiator for the formation of a polyepoxide network if the network is to be formed at room temperature while being exposed to light or other radiation (fig. 5) [8]. We have studied this network formation and the microstructure of the net-

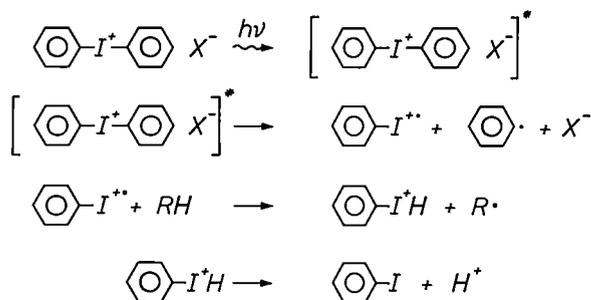


Fig. 5. The photochemical formation of an acid from a diaryl-iodonium salt [8], where X⁻ represents AsF₆⁻, SbF₆⁻, etc. In the first reaction step the diaryl-iodonium salt is brought into an excited state (*), which then decays into reactive particles, mainly cations (+), radicals (•) and radical cations (+•).

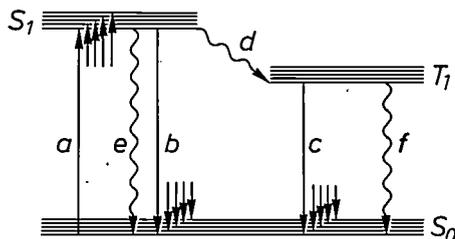


Fig. 6. Simplified representation of the Jablonski diagram for S₀, S₁ and T₁ of an organic compound. Absorption (a), fluorescence (b), phosphorescence (c), intersystem crossing (d) and non-radiative transitions (e and f) are represented. The small arrows indicate the fine structure of the vibration.

works formed with the aid of phosphorescence spectroscopy, a method that has previously only been used in research on linear polymers.

Since most organic compounds possess a singlet ground state S₀, the absorption and emission of light by these compounds can be represented in a simplified Jablonski diagram (fig. 6). In the case of phosphorescence the radiative transition takes place from an excited relatively long-lived triplet state (T₁). Its long lifetime makes this triplet state extremely sensitive to non-radiative decay, which can occur as a result of molecular vibrations or collisions with other molecules in a triplet state (e.g. oxygen molecules). To record phosphorescence spectra it is therefore always necessary to limit the molecular mobility as far as possible. One way of doing this is by performing the measurements at low temperatures on an amorphous glass in which the molecule of interest has been dissolved.

As noted in the introduction, during polymer formation the product gradually becomes more viscous. This means that its molecules become less mobile. In the cured network the molecular mobility is low but is still subject to temperature-dependent changes. In this stage the lifetime of the triplet state and the intensity of the phosphorescence can provide important information about dynamic changes in the network.

The half-life τ of the triplet state is given by:

$$\tau^{-1} = \tau_r^{-1} + \tau_{nr}^{-1},$$

where τ_r is the lifetime of the triplet state in the radiative decay and τ_{nr} is its lifetime in the non-radiative decay. The temperature dependences of intensity and

[6] P. J. Flory, Principles of Polymer Chemistry, Cornell University Press, Ithaca, N.Y., 1953, pp. 69-104.

[7] The following books should interest those who would like to know more about this subject;

on polymer chemistry:

G. Odian, Principles of Polymerization, 2nd edition, McGraw-Hill, New York 1981;

F. W. Billmeyer, Jr., Textbook of Polymer Science, Wiley Interscience, New York 1971;

D. W. van Krevelen, Properties of Polymers, Elsevier, Amsterdam 1976;

P. J. Flory, Principles of Polymer Chemistry, Cornell University Press, Ithaca, N.Y., 1953;

on photochemistry and photophysics:

J. E. Guillet, Polymer Photophysics and Photochemistry, Cambridge University Press, Cambridge 1985;

N. J. Turro, Modern Molecular Photochemistry, Benjamin/Cummings, Menlo Park, Cal., 1978;

on nuclear magnetic resonance spectroscopy for organic molecules:

H. Günther, NMR-Spektroskopie, Thieme Verlag, Stuttgart 1973;

J. B. Stothers, Carbon-13 NMR Spectroscopy, Academic Press, New York 1972.

[7] E. W. Meijer and R. J. M. Zwieters, in: Crosslinked Epoxies, B. Sedláček and J. Kahovec (eds), Walter de Gruyter, Berlin 1987, pp. 27-40.

[8] J. V. Crivello, Cationic polymerization — iodonium and sulfonium salt photoinitiators, Adv. Polym. Sci. 62, 1-48, 1984.

half-life are well known from work on linear polymers, and are given by^[9]:

$$(I_f/I_p)(I_{p,0}/I_{f,0}) - 1 = A \exp(-E_a'/RT),$$

$$\tau^{-1} - \tau_0^{-1} = B \exp(-E_a''/RT),$$

where I_f , $I_{f,0}$, I_p and $I_{p,0}$ are respectively the intensities of the fluorescence at the temperatures T and $T = 0$, and of the phosphorescence at the temperatures T and $T = 0$. τ and τ_0 are the lifetimes at T and $T = 0$. A and B are constants; E_a' is the energy term denoting the temperature-dependence of the intensity of the phosphorescence and E_a'' is the energy term denoting the temperature dependence of its lifetime.

Fig. 7 shows emission characteristics of a polyepoxide network that we have studied. Intermediates that occur during the photoinitiated decomposition of the diaryl-iodonium compounds used as a latent cat-

alyst are found to give rise to fluorescence with a maximum at a wavelength of 340 nm and to phosphorescence with a maximum at a wavelength of 480 nm (fig. 7a). The phosphorescence decays almost exponentially with a half-life of 1.6 s at room temperature (fig. 7b). In our investigations we used the diaryl-iodonium compound di(4-*tert*-butylphenyl)-iodonium-hexafluoroarsenate as latent catalyst and emission source. Although the nature of the emitting species has not been identified in detail, this is no objection to using the compound as a 'probe'.

Molecular mobility as a function of polymerization time^[10]

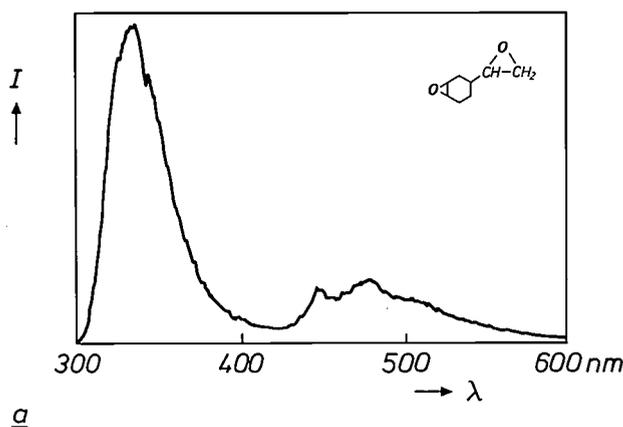
If the fluorescence and phosphorescence intensities are measured as a function of the time in which the polymerization and cross-linking take place, an exponential increase is observed in both cases. The immediate rise in intensity of the fluorescence at the start of polymerization is followed, after an induction period dictated by the monomer used, by a rise in the intensity of the phosphorescence. In the case of the phosphorescence the induction period and the slope of the increase in intensity are interrelated (fig. 8).

As stated, all these effects are determined by the rate at which the molecular mobility in the continuously cross-linking polymer matrix is slowed down. This slowing down of the mobility is determined by the reaction rate of the polymerization and the change in viscosity on a micro-scale brought about by each step in the polymerization. It is the combination of these factors that interests the polymer chemist and can be investigated with this method of phosphorescence spectroscopy.

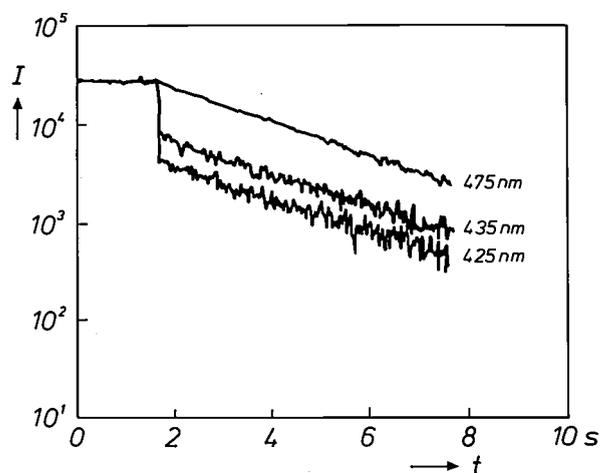
The end of the induction period indicates that the molecular mobility has reached a certain threshold value. The slope of the increase in phosphorescence is a measure of further changes in mobility.

Small variations in the composition of the original materials are found to give rise to changes in the polymerization processes that can readily be measured. Fig. 9 shows the result of phosphorescence measurements on mixtures of 1-epoxyethyl-3,4-epoxycyclohexane and 1,4-butanediol-diglycidyl ether. The first bisepoxide has a rigid structure and reacts rapidly, whereas the second is flexible and reacts relatively slowly. This has the result that when the proportion of the flexible monomer in the original mixture increases, it takes longer for the molecular mobility to slow down sufficiently for phosphorescence to occur. This is reflected in a longer induction period and not so steep a rise in the intensity curve.

We have compared the results of this new method of studying the processes of polymerization and net-



a



b

Fig. 7. a) The emission spectrum (intensity I against wavelength λ) of a network of 1-epoxyethyl-3,4-epoxycyclohexane (fig. 3c), recorded at an excitation wavelength of 254 nm at room temperature. b) The decay of the excited triplet state for the same network, detected at three different wavelengths indicated in the figure. The excitation wavelength was 290 nm.

work formation with results of microcalorimetric experiments (DSC for Differential Scanning Calorimetry) on the same epoxides. In these experiments we monitored the progress of the reaction by measuring the heat of reaction produced per unit time. A com-

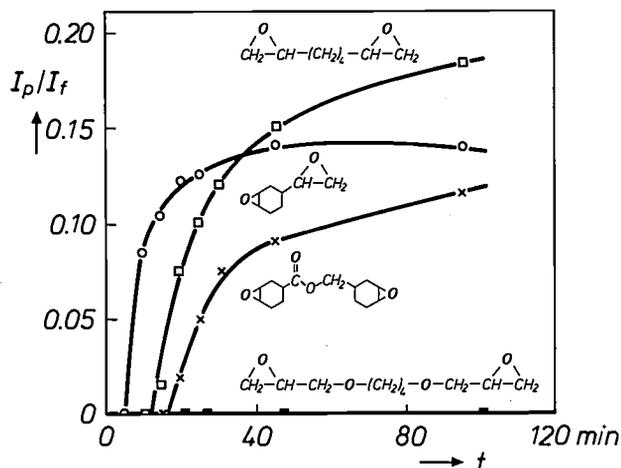


Fig. 8. Relative increase I_p/I_f in the phosphorescence intensity as a function of the polymerization time t for four polyepoxides at room temperature.

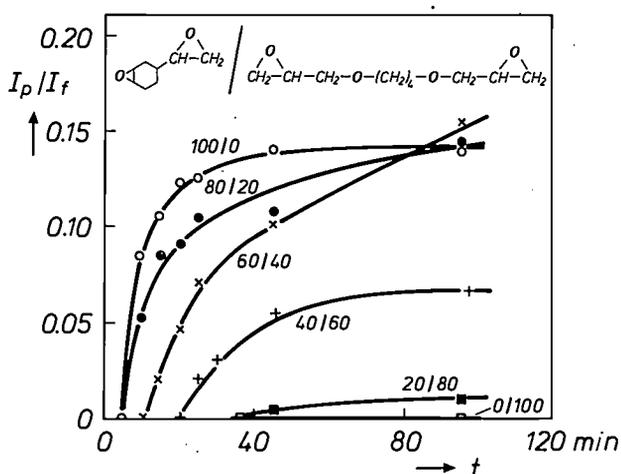


Fig. 9. As in fig. 8., for mixtures of two epoxides.

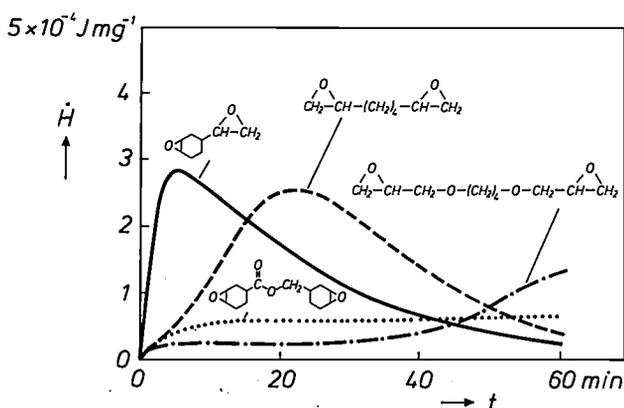


Fig. 10. Result of microcalorimetric (DSC) measurements on the photopolymerization of four bisepoxides at room temperature. The heat flux \dot{H} is plotted as a function of the reaction time t .

parison of fig. 8 and fig. 10 shows that both methods lead to corresponding conclusions: listing the various bisepoxides in order of polymerization rate as measured with the microcalorimeter gives exactly the same sequence as listing them in order of network formation as determined by phosphorescence spectroscopy.

The phosphorescence measurements had therefore taught us something about the rate of polymerization, but the method still had more to offer.

We also wanted to monitor the progress of the polymerization from measurements of the half-life of the excited triplet state. To determine this half-life we made a 'time-resolved' measurement of the phosphorescence. It was then found that the *half-life* of the triplet state is *not* dependent on the polymerization time, unlike the *intensity* of the phosphorescence, which does depend on the polymerization time. We assume that the explanation for this remarkable effect must reside in the inhomogeneous structural formation of the polyepoxide network. It seems to us that the increase in the degree of cross-linking does not take place homogeneously but is localized. This is to say that during the formation of the network the matrix consists of areas with a relatively high degree of conversion (giving phosphorescence with a long half-life) and areas with a relatively low degree of conversion (no phosphorescence). Extension of the regions with a high degree of conversion at the expense of areas with a low degree of conversion will then *increase* the intensity of the phosphorescence, but will not affect the half-life.

Other indications of the inhomogeneous polymerization postulated here have been found from electron-resonance-spectroscopy measurements on trapped free radicals in these polyepoxides^[10]. Inhomogeneous polymerizations have also been postulated in other polymer networks^[11].

Molecular mobility as a function of temperature^[12]

As well as information about the rate of polymerization, phosphorescence spectroscopy can also provide information about dynamic-structural as-

[9] A. C. Somersall, E. Dan and J. E. Guillet, Photochemistry of ketone polymers XI, *Macromolecules* 7, 233-244, 1974;

K. J. Smit, R. Sakurovs and K. P. Chigginio, Temperature dependence of fluorescence and phosphorescence from probe molecules in polymer substrates, *Eur. Polym. J.* 19, 49-53, 1983.

[10] E. W. Meijer, D. M. de Leeuw, F. J. A. M. Greidanus and R. J. M. Zwiers, Molecular mobility studies in polyepoxides obtained by photoinitiated cationic polymerization, *Polym. Commun.* 26, 45-47, 1985.

[11] J. G. Kloosterboer, Network formation by chain crosslinking photopolymerization and its applications in electronics, *Adv. Polym. Sci.* 84, 1-61, 1988.

[12] E. W. Meijer and R. J. M. Zwiers, Molecular mobility of polyepoxide networks as revealed by emission spectroscopy and mechanical analysis, *Macromolecules* 20, 332-338, 1987.

pects of the three-dimensional network ultimately formed. To obtain this information we measure the intensity of the phosphorescence and the lifetime of the triplet state as a function of temperature. This is illustrated by the results of measurements on the relatively flexible network of 1,4-butanediol-diglycidyl ether. At room temperature it will not normally be possible to produce phosphorescence from such a network, but by reducing the temperature the molecular mobility of the network can be reduced sufficiently for phosphorescence to occur (fig. 11).

The increase in the intensity of the phosphorescence and the lifetime of the associated triplet state when the temperature is reduced can most easily be investigated by plotting the measured values as Arrhenius curves (fig. 12). Both curves in the figure consist of

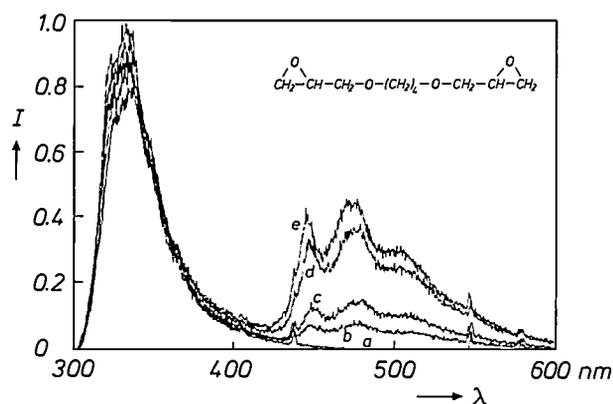


Fig. 11. Emission spectra of a network of 1,4-butanedioldiglycidyl ether (fig. 3d) for different temperatures at an excitation wavelength of 254 nm: a) 296 K, b) 279 K, c) 273 K, d) 233 K, e) 193 K.

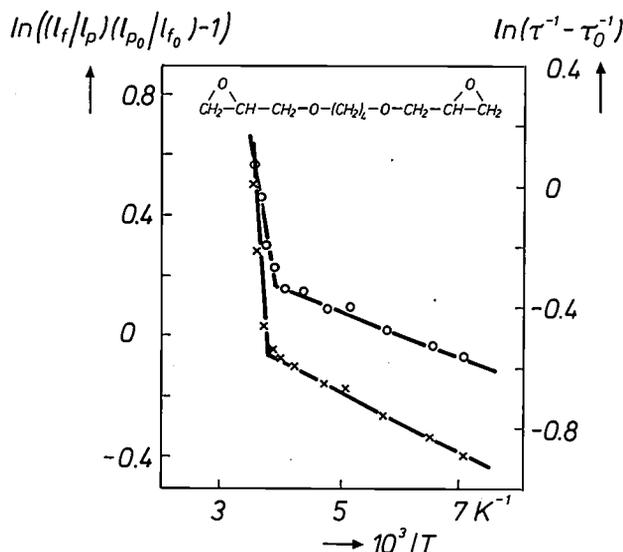


Fig. 12. Temperature dependence of the (relative) phosphorescence intensity (left) and of the (relative) half-life of the excited triplet state for a network of 1,4-butanedioldiglycidyl ether (fig. 3d).

two linear sections with a sharp transition at a temperature of about 260 K. The slope of a straight section is a measure of the activation energy of the mobility (or microviscosity) in the network. The low-temperature section (with $E_a = 6$ to 8 kJ/mol) corresponds to a state of low mobility and therefore little non-radiative decay from the triplet state. The changes are mainly attributable to molecular vibrations and to local degrees of freedom in conformation. The high-temperature section corresponds to a state of greater mobility ($E_a \geq 150$ kJ/mol) and the radiative decay and the non-radiative decay of the triplet state take place at a comparable rate. Here *parts* of the macromolecule are able to move, culminating eventually in the macroscopic glass-rubber transition.

These Arrhenius curves are similar for all the polyepoxide networks that we have studied (Table I). The slopes of all the low-temperature sections are the same, and the transition to the high-temperature section is at about 260 K on all the curves; only the slopes of the high-temperature sections differ. It is only in these slopes that the considerable differences between the molecular structures of the initial monomers become apparent. And indeed they show that the mobilities of the networks built up from flexible molecules like 1,4-butanediol-diglycidyl ether have a higher activation energy than those of the networks built up from rigid molecules like 1-epoxyethyl-3,4-epoxycyclohexane.

When the polyepoxides mentioned above are subjected to DMTA measurements (dynamic mechanical thermal analysis) it is found that the polymers built up from flexible monomers have a narrow transition region at low temperatures, and that the polymers built up from rigid molecules have a broad transition region at higher temperatures.

A comparison of the phosphorescence measurements and the mechanical measurements reveals the following picture of the dynamic structure of polyepoxide networks. In all the polyepoxides that we have studied the molecular mobility at very low temperatures is the same below the threshold value. In all the polyepoxides the molecular mobility starts to change at the same temperature, about 260 K. The movements that then occur are related to the molecular mobility of the individual groups. In this part of the curve above 260 K we therefore see a premonition of the macroscopic glass-rubber transition state, which will appear for the most mobile network at the lowest temperature when the temperature is increased again.

This — non-destructive — method of phosphorescence spectroscopy therefore gives us valuable supplementary information about the microscopic structure of polymer networks and the complicated way in which they are formed.

Table I. Experimentally obtained values for the activation energies, determined from Arrhenius curves as shown in fig. 12. The columns E_{a1} give the values in kJ/mol as determined for the low temperatures, and columns E_{a2} give the values for the higher temperatures. The columns T give the transition temperature in K.

Polymer networks of the epoxides in fig. 3	Intensity			Half-life		
	E_{a1}	T	E_{a2}	E_{a1}	T	E_{a2}
<i>c</i>	8	259	14	3	240-250	9
<i>d</i>	8	262	170	6	263	115
<i>c:d</i> = 4:6	8	266	22	5	263	17
<i>b</i>	6	225-257	23	5	245-250	10
<i>b:e</i> = 1:1	4	259	42	3	259	14

Sometimes it will also be desirable to find out more about the particular local reaction mechanism for each polymerization step during the formation of a network. An investigation of this type that we performed with the aid of NMR spectroscopy is described in the next section.

A mechanism for epoxide polymerization

An investigation in which the mechanism of a promising new catalyst-initiator system for epoxide polymerization was studied^[13] will now be described as our second example.

The high thermal and chemical stability of polyepoxide networks, combined with good electrical insulation, makes polyepoxides eminently suitable as encapsulating materials for electrical components. If these networks are to be produced by an ionic reaction mechanism, generally a very effective method, then it is particularly important that no ionogenic impurities such as catalysts should be left behind in the network. A likely material investigated for this purpose was tris(acetylacetonato) aluminium — referred to from now on as $Al(acac)_3$ — in combination with phenolic compounds. Nothing was previously known about the mechanism of the polymerization produced with this system.

NMR spectroscopy as an investigative method

The theoretical background of the nuclear-magnetic-resonance spectroscopy we used for this investigation has been dealt with several times in this journal^[14], but with no particular attention to the opportunities offered by NMR spectroscopy for a better understanding of the structure of complicated organic compounds (including polymers). This form of spectroscopy can be used with particular advantage for studying the hydrogen atoms (with spin quantum number 1/2) in these compounds.

An external magnetic field causes the energy associated with a nuclear spin to 'degenerate' into two

energy levels, one by aligning the nuclear spin with the field, and the other by aligning it in the opposite sense to the field. A resonance between these two levels gives rise to NMR signals. Because there are differences in the local magnetic field around a nuclear spin, due to the electron distribution in the molecule, the energy required for the resonance of each physically non-identical hydrogen atom does not have exactly the same value. These differences in resonance energy can be observed via the 'chemical shift' in the NMR spectrum, which is characteristic of the various substitution patterns in the molecule.

The spin-spin couplings that frequently occur have the effect of splitting the resonances, and are therefore an additional source of information. Spin-spin couplings in NMR spectroscopy arise from the splitting of energy levels of nuclear spins caused by the presence of adjacent spins (at a distance of up to about three covalent bonds). The parallel and antiparallel combinations cause the splitting. Identical spins do not split one another.

NMR spectroscopy can also identify other nuclei in much the same way as those of the hydrogen atoms. In organic chemistry NMR spectroscopy is very useful for studying ^{13}C atoms. The use of powerful magnetic fields (up to 13.8 Tesla), with high-resolution equipment and the application of various complicated pulse techniques, have meant that NMR spectroscopy has now become one of the most powerful spectroscopic methods for analysing organic molecules. The impressive results reported have included the explanation of the structure of extremely complex proteins and parts of DNA^[15]. Polymers have also been the subject of extensive investigations with NMR^[16].

Most information is now obtained from liquid systems, although solid-state NMR is in full development. We have investigated the mechanism of chain polymerization with the aid of a model compound, a monoepoxide. This forms soluble polymers, making high-resolution NMR possible.

[13] Contributions to this part of the investigation were made by F. W. v.d. Wey, J. P. H. Heynen, H. Draayer, A. W. de Bruyn, A. Schokker and C. Kruk.

[14] P. R. Locher, Proton NMR tomography, Philips Tech. Rev. 41, 73-88, 1983/84.
D. J. Kroon, Nuclear magnetic resonance, Philips Tech. Rev. 21, 286-299, 1959/60.

[15] R. Kaptein, S. Stob, R. M. Scheek, K. Dijkstra and T. W. Schleich, High-field and photo-CIDNP studies of biological molecules, Bull. Magn. Reson. 2, 28-31, 1981;
G. Wagner and K. Wüthrich, Observation of internal mobility of proteins by nuclear magnetic resonance in solution, Methods Enzymol. 131, 307-328, 1986.

[16] J. R. Havens and J. L. Koenig, Applications of high-resolution carbon-13 nuclear magnetic resonance spectroscopy to solid polymers, Appl. Spectrosc. 37, 226-249, 1983;
F. A. Bovey, New directions in the nuclear magnetic resonance spectroscopy of polymers, Polym. Eng. & Sci. 26, 1419-1428, 1986.

The initiator-catalyst system^[17]

As demonstrated elsewhere, many metal-acetyl acetonates, in combination with very different phenolic compounds, can be used to bring about the polymerization of epoxides, but the most effective one is the combination of $\text{Al}(\text{acac})_3$ with catechol (*ortho*-dihydroxybenzene)^[18]. When this combination reacts with cyclohexene oxide (as a 'model monomer'; see fig. 3), a polymer is formed as shown in fig. 13. As can be seen, catechol acts as the initiator, while $\text{Al}(\text{acac})_3$ is the catalyst.

During the polymerization an intermediate can be observed by $^1\text{H-NMR}$ spectroscopy. This same intermediate also arises when $\text{Al}(\text{acac})_3$ and catechol react with each other *in the absence of epoxides*. Closer study of the reaction product reveals that a dimer structure is formed. This is preceded by a 'ligand' exchange, in which an acetyl-acetone group makes way for a catechol group^[17] (fig. 14).

Because of this dimerization an intramolecular hydrogen bridge can form between the two phenolic groups present in the dimer, resulting in a strongly 'acidic' proton. As we have seen earlier, an acid can catalyse the polymerization of epoxides.

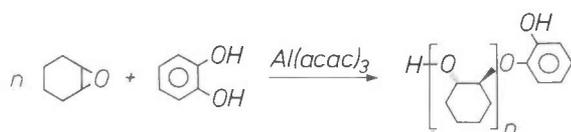


Fig. 13. The polymerization of cyclohexene oxide (fig. 3e) using the combination of $\text{Al}(\text{acac})_3$ and catechol.

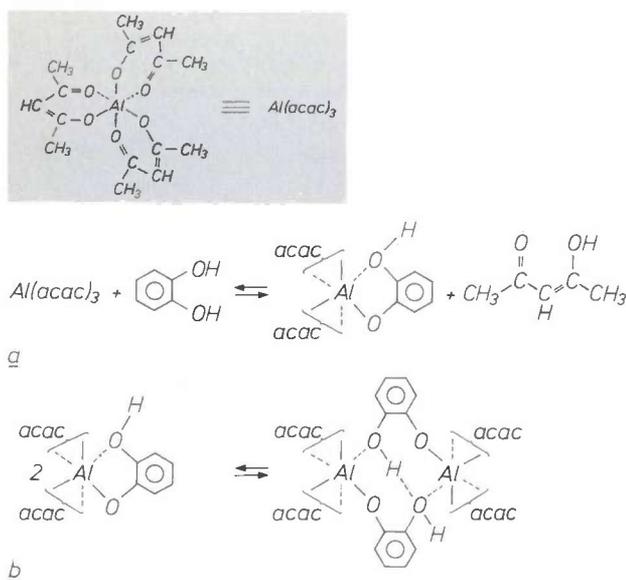


Fig. 14. Reactions for ligand exchange between $\text{Al}(\text{acac})_3$ and catechol (a) and for dimerization (b). See also fig. 15.

The dimer consists of a complex combination of three geometrical isomers. In solution these isomers are in equilibrium with each other, and there is also a monomer-dimer equilibrium (see fig. 15). Both equilibria have been investigated in detail by NMR spectroscopy^[17].

The mechanism of linear-chain polymerization

As we have seen, dimer formation is essential for effective catalysis and initiation. Two important reaction steps are evidently necessary for polymerization. In the first place the catechol must be activated to open the first epoxide ring; this is done by the ligand exchange with $\text{Al}(\text{acac})_3$. In the second step this ligand exchange results in the formation of the dimer; this is the 'acid' dimer that acts as the actual catalyst in the polymerization. Since the re-formed catalyst $\text{Al}(\text{acac})_3$ is normally an electrically neutral non-ionic compound, it does indeed seem possible to obtain polymerization in this way without a residue of ionic 'contamination'.

By using chromatography and titration we were able to demonstrate that the formation of the dimer is very rapid, in appropriate conditions, and that the propagation step of the polymerization is the rate-determining step (fig. 16). It can also be deduced from the figure that the polymerization is just as fast with the combination of $\text{Al}(\text{acac})_3$ and catechol as it is with a dimer obtained by separate synthesis. The figure also shows that the polymerization stops when the initiator catechol is used up. This is because the absence of the initiator makes further dimer formation impossible, so that both initiation and propagation of the polymerization come to a stop.

The need for the presence of an acid catalytic dimer suggests that the polymerization takes place via a cationic mechanism (and not for example via a coordinative mechanism, in which an epoxide is repeatedly inserted between the aluminium atom and the growing chain).

In an attempt to obtain further evidence of a cationic mechanism in the epoxide polymerization that we were studying, we investigated a special epoxide monomer, *exo*-2,3-epoxy-norbornane (structure *a* in fig. 17), which served as a model. In a series of experiments we tried to polymerize this compound by using a number of different catalysts. We found that polymerization was not produced by anionic or coordinative catalysis, but only when the catalysis contained a cationic mechanism. The observation that *exo*-2,3-epoxy-norbornane also polymerizes with a combination of $\text{Al}(\text{acac})_3$ and catechol, or with the dimer, is therefore another indication of the cationic nature of this system.

Characterization of the polymer obtained from the acid-catalysed polymerization of *exo*-2,3-epoxy-norbornane revealed that a Wagner-Meerwein rearrangement takes place during this polymerization. In such a rearrangement the normal 2,3-substitution (structure *b* in fig. 17) does not occur in the polymer produced, but a 2,7 substituted product is obtained instead (structure *c* in fig. 17).

To prove that such an abnormal substitution took place, we subjected the polymer formed to a number of two-dimensional spin-spin decoupling correlated-spectroscopy (COSY) experiments^[19] (fig. 18), in which we compared the spectrum of the polymer with that of low-molecular compounds like 2,7-dihydroxy-norbornane.

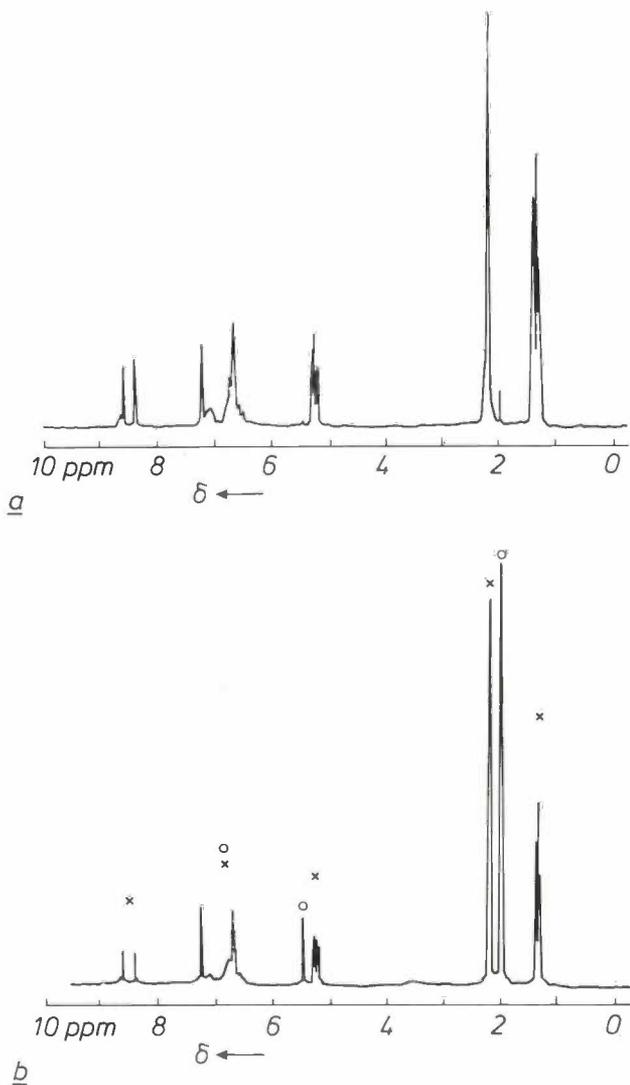


Fig. 15. a) ¹H-NMR spectrum of a freshly made solution of the dimer in CDCl₃. Note the appearance of different resonances for the acetyl-acetate groups at $\delta = 1.2$ and 5.2 ppm (parts per million) and the three 'acid' protons at $\delta = 8-9$ ppm. b) ¹H-NMR spectrum of a solution of the dimer in CDCl₃ after 2 hours at room temperature. The resonances indicated by an asterisk * correspond to the dimer, and those indicated by o to the monomer.

Polyepoxide network formation

In further investigation of the network formation of bisepoxides initiated and catalysed by the combination of catechol and Al(acac)₃, we encountered an

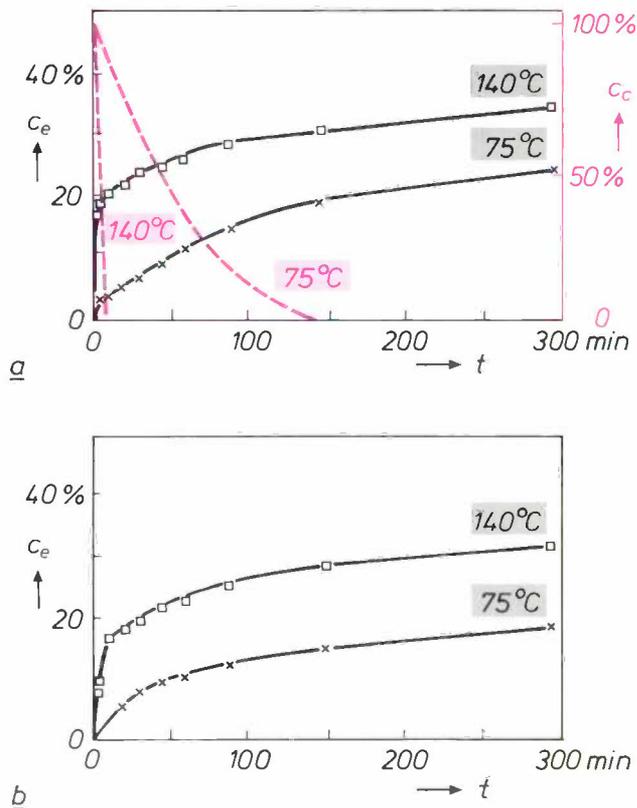


Fig. 16. The percentage c_e of the epoxide conversion for cyclohexene oxide with the combination of Al(acac)₃ and catechol (a) and with dimer (b) plotted as a function of the reaction time t of polymerization at $T = 140^\circ\text{C}$ and 75°C . The concentration of catechol c_c as a function of the reaction time is also shown in 16a.

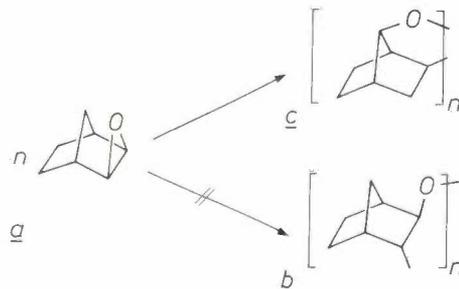


Fig. 17. Structure of *exo*-2,3-epoxy norbornane (a; see also fig. 3f), a 2,3-substituted norbornane (b) and a 2,7-substituted norbornane (c).

[17] E. W. Meijer, The synthesis and structure of bis(acetylacetonato) (2-hydroxyphenolato) aluminium: a stable dimer, *Polyhedron* 6, 525-529, 1987.
 [18] M. Markovitz, Controllable reactivity epoxy resin hardeners, in: *Chemistry and Properties of Crosslinked Polymers*, S. S. Labana (ed.), Academic Press, New York 1977, pp. 49-58.
 [19] R. Benn and H. Günther, Moderne Pulsfolgen in der hochauflösenden NMR-Spektroskopie, *Angew. Chem.* 95, 381-411, 1983.

additional complicating factor: the dimeric initiator-catalyst was not always soluble in the bisepoxides selected. We found, however, that this complication could also be turned to good use.

One of the bisepoxides that we studied was (3,4-epoxycyclohexyl)methyl-3',4'-epoxycyclohexane carboxylate (see fig. 3). We polymerized this compound under various reaction conditions. The reactions were monitored by titration and by determining the gel point.

From this investigation we drew the following conclusions:

- The dimer in the dissolved form is the best initiator-catalyst.
- If the dimer is added in the solid form, the polymerization does not take place or only very slowly at low temperatures, because the dimer dissolves only slightly.
- At higher temperatures fast polymerization and cross-linking do take place with a solid dimer, though

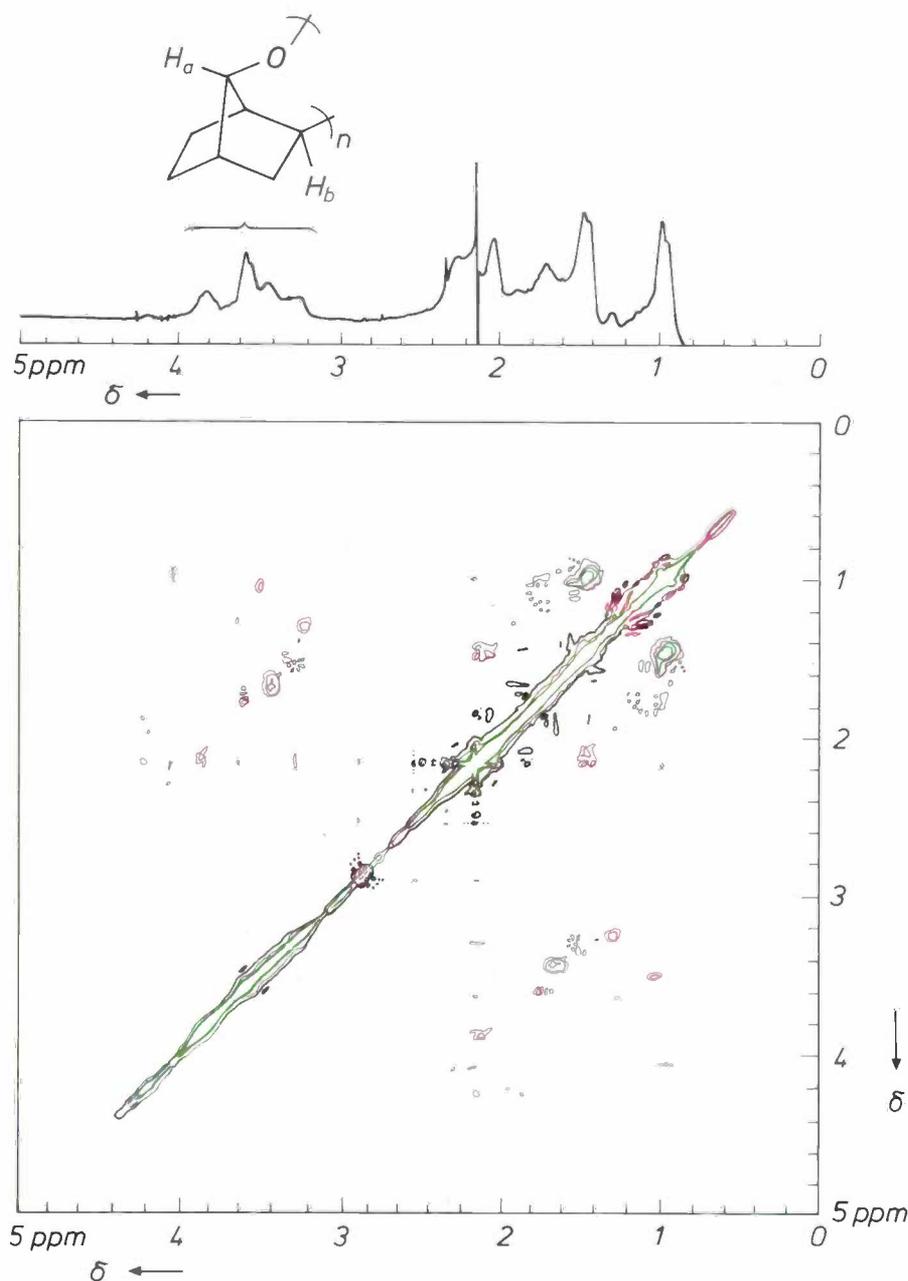


Fig. 18. $^1\text{H-NMR}$ and COSY spectra of the polymer of exo-2,3-epoxy-norbornane. COSY experiments (CORrelated SpectroscopY) indicate whether there is a spin-spin coupling between two hydrogen atoms. This coupling appears as a cross peak in the two-dimensional spectrum, where the normal $^1\text{H-NMR}$ spectrum appears as a diagonal in the contour picture. The absence of spin-spin coupling between the hydrogen atoms H_a and H_b in the polymer (between $\delta = 3.1$ and 4.0 ppm) is a strong indication that these atoms are in a 2,7 and not in a 2,3 relationship.

only after some delay. Initially, a reaction is only possible at the interface between the solid dimer and the liquid bisepoxide. The dimer is more soluble in polymerized bisepoxide, so that the rate of polymerization increases during the reaction.

- The temperature dependence of the rate of polymerization is not described by a simple Arrhenius equation, because of an irreversible change in the solid dimer [17].
- In the case of the catechol/ $\text{Al}(\text{acac})_3$ combination we have a situation intermediate between the two cases mentioned above (fig. 19).

The above conclusions show that the addition of the dimer in solid form produces a latent initiator-catalyst system.

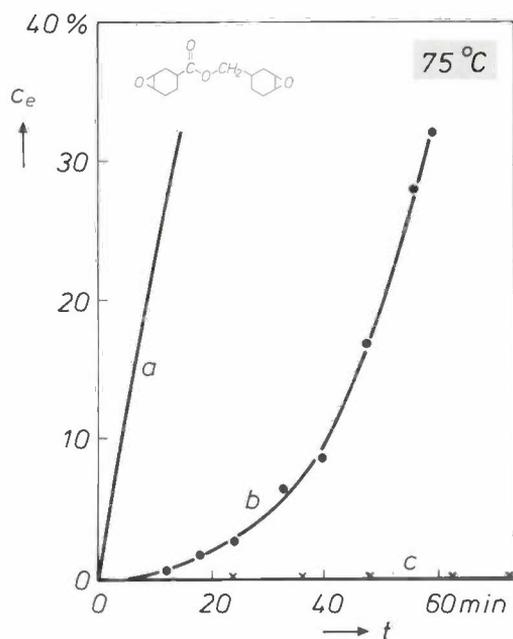


Fig. 19. The epoxide conversion (c_e) as a function of polymerization time t , as measured by means of titration. It is found to make a great difference whether the dimer is added (a) in a solution (of CHCl_3) or (c) as a solid. b) Progress of the epoxide conversion with the combination of $\text{Al}(\text{acac})_3$ and catechol.

A polymer network that can be built up from such a two-phase situation has the advantage of remaining stable for a long time at low temperatures; fast polymerization and cross-linking can only take place at high temperatures, because a homogeneous reaction mixture is then formed.

Good use can be made of this form of latent catalysis in the production of polymer networks, if at least the perfectly homogeneous phase is reached. If the solid phase is excessively reactive and insufficiently soluble in the reaction mixture, the inevitable result will be an inhomogeneous polymer network. The associated problems are still the subject of investigation.

The two investigations described have been chosen to give some idea of the approach now possible in this sector of chemistry. One shows how phosphorescence spectroscopy can be used to find the relationship between the molecular mobility and the physical properties, and the other how nuclear-magnetic-resonance spectroscopy can help in sorting out the highly complex reaction mechanisms. Both have given a better understanding of the relation between the structure and properties of polyepoxides.

Summary. The polymerization of epoxide monomers with catalysts or initiators results in polyepoxides. Special attention is given to chain polymerization from bisepoxides: the three-dimensional networks formed permit a wide variety of applications. Methods of gaining a fundamental understanding of the formation of these networks and their structure are illustrated by two examples. In the first example it is shown how phosphorescence spectroscopy can be used to monitor changes in the molecular mobility of polymer segments, both during the formation of the network and as a function of temperature. In the second example it is shown how nuclear-magnetic-resonance spectroscopy can be used to study the mechanism of network formation with the aid of model compounds. Emphasis is placed here on the part played by the $\text{Al}(\text{acac})_3$ /catechol initiator-catalyst system.

1938

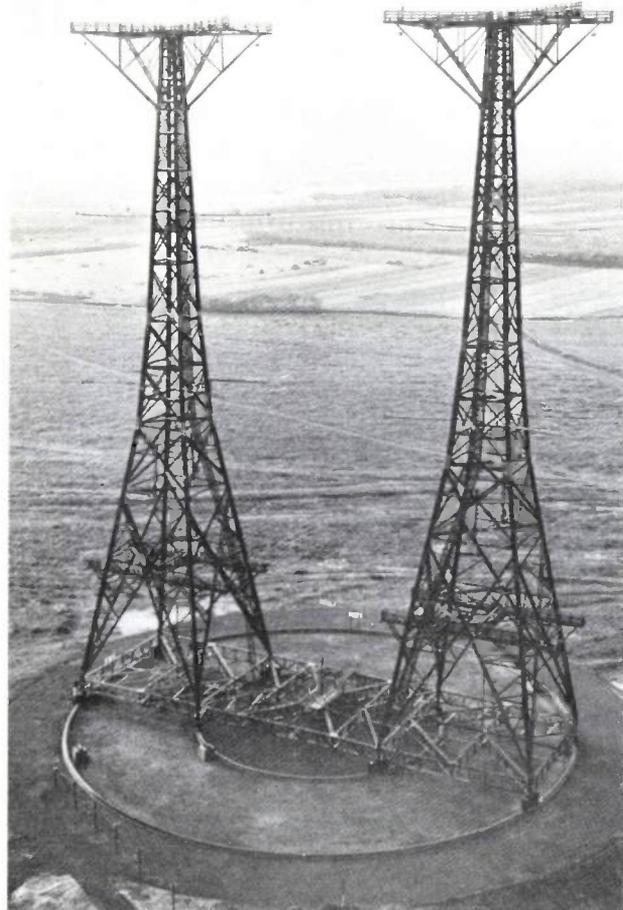
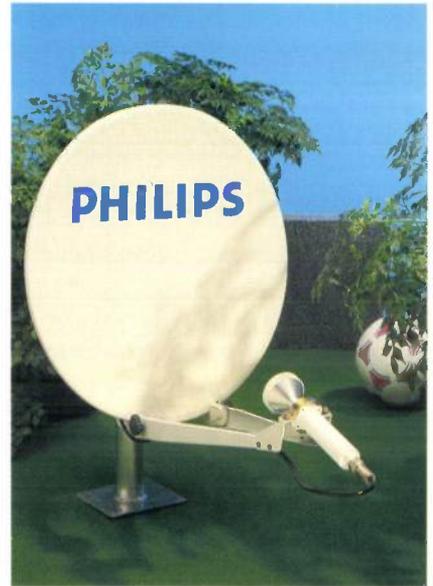
THEN AND NOW

1988

Antennas

Antennas are the 'mouths and ears' of many electrical communications systems. They very largely determine the distances over which radio signals can be transmitted and the frequencies that can be used. Their shape and form is of almost limitless variety and depends on the particular application. Half a century ago short-wave radio broadcasting was an important means of communication between the various continents. A special antenna for transmissions on about 10 MHz (a wavelength of about 30 m) was built in the Netherlands. This consisted of two wooden towers, 60 metres high, with a number of vertical half-wave antenna wires attached to them (black-and-white photograph) [1]. The entire structure could be rotated and was mounted on a circular rail system with an outer diameter of just over 46 metres.

In 1988 we have very many more broadcasting systems. In one of the newest ventures television signals are directly transmitted via a geostationary satellite that stays apparently stationary at a height of about 36 000 km (22 500 miles) above the earth (Direct Broadcasting Satellite). Special antennas give direct individual reception from the satellite. Since the frequencies used are higher than 10 GHz the antennas



for transmission and reception are not made from wires, but are parabolic dishes. These act as reflectors for the actual transmitting or receiving element located at the focus. The colour photograph shows a modern version of a satellite receiver antenna with an offset feed. This is designed to bring the focus close to the bottom of the dish, so that the almost vertical antenna is directed towards the satellite. Because of the high quality of the receiving equipment this antenna need only have a diameter of about 55 cm (22 inches). With a satellite, one or more television programmes can be transmitted to an entire continent. It is also possible to take advantage of the full potential of the new and improved television standards such as D2-MAC and HD-MAC [2].

[1] From Philips Technical Review, February 1938.

[2] See for example: M. J. J. C. Annegarn *et al.*, HD-MAC: a step forward in the evolution of television technology, Philips Tech. Rev. 43, 197-212, 1987.

Theory and practice of acoustic noise control in electrical appliances

J. Crucq

'Life is full of sounds and we want to hear the pleasant and vital ones, while shunning the unpleasant and dangerous variety.' So wrote R. B. Lindsay in his historical introduction to a re-issue of Lord Rayleigh's famous standard work 'The Theory of Sound'^[*]. *How appropriate that today, with the Compact Disc delighting the ear with agreeable sound, there is renewed interest in the suppression of sound that is less welcome. Noise suppression has become an increasingly important aspect of the design of electrical appliances. It has provided some of the motivation for studies in recent years at Philips Research Laboratories in Eindhoven that are yielding a more fundamental understanding of the generation and suppression of unwanted sound.*

Introduction

In nearly all electrical equipment and appliances — shaver, Compact Disc player, television set or X-ray diagnostic machine — some attention must be given to noise suppression at the design stage. Noise suppression must be based on a fundamental understanding of the physical mechanisms responsible for the origin and transmission of sound^[1]. When a good understanding of the fundamentals is available, there is often no need to resort to an empirical approach, which usually amounts to taking a number of steps to reduce noise and measuring the effect of each step separately.

Noise suppression for an appliance still on the drawing board is not usually realistic. What is generally done, therefore, is first to identify the different sources of acoustic noise in a prototype or production model. The next step is to establish the sound power level of the individual sources, usually in dB with respect to a level of 10^{-12} W. The results should then show which sources contribute most to the total noise produced. The investigation should be aimed at these sources first of all. This involves characterizing the

sources by the frequencies of the constituent components of the noise signal and by their behaviour as a function of the load and other operating conditions.

The process of transferring the sound from a given source to the environment contains a number of steps: excitation, transmission and acoustic radiation. Excitation is the mechanism that produces vibration. In an electric motor, for example, vibration is due to rotor imbalance or the carbon brushes striking the segments of the commutator. Transmission is the transfer of vibration from the source to other parts of the mechanical structure and therefore also to the outside of the appliance. Acoustic radiation is the conversion of vibration at the exterior of the appliance into airborne sound waves.

The relationship between excitation, transmission and radiation can be shown for each link in the

[*] J. W. Strutt, 3rd Baron Rayleigh, *The theory of sound*, parts I and II, Dover, New York 1945.

[1] P. M. Morse, *Vibration and sound*, Am. Inst. of Physics, 1976;
 J. P. den Hartog, *Mechanical vibrations*, McGraw-Hill, New York 1956;
 W. C. Hurty and M. F. Rubinstein, *Dynamics of structures*, Prentice-Hall, Englewood Cliffs, N.J., 1964;
 L. L. Beranek, *Acoustics*, McGraw-Hill, New York 1954.

sound-transfer chain by writing an equation based on an analogy with electrical networks:

$$F = vZ, \quad (1)$$

where F is the force associated with the vibration, v the velocity, and Z the mechanical impedance; all these quantities are complex. F is equivalent to the voltage and v to the current in this 'mechanical Ohm's law'.

If we only consider r.m.s. values of F and v [2], then for the excitation we have:

$$F_{1,rms} = v_{1,rms} |Z_1|, \quad (2)$$

where $v_{1,rms}$ is the r.m.s. value of the velocity of the vibration at the point of action of the alternating force of r.m.s. value $F_{1,rms}$, and Z_1 is the mechanical input impedance of the structure at this position. The quantities in this and the following equations are thus all position-dependent and are also functions of the frequency f of the vibration. For the propagation of the vibrations through the appliance we have:

$$v_{2,rms} = H v_{1,rms}, \quad (3)$$

where H is the real, averaged transfer function of the structure and $v_{2,rms}$ is the r.m.s. value in time and position of the velocity of the vibration at the cabinet or casing of the appliance. The power P of the sound waves that are 'radiated' from the casing into the surrounding air is then given by:

$$P = \text{Re}(Z_2) v_{2,rms}^2, \quad (4)$$

where Z_2 is the mechanical radiation impedance of the casing.

The mechanical radiation impedance is thus defined as the complex ratio of force to velocity (eq. 1). Another relevant quantity, the acoustic radiation impedance, is defined as the ratio of the pressure to the volume flow of the gas in which the radiated vibration propagates. The mechanical and acoustic radiation impedances are mathematically related through the ratio $S^2:1$, where S is the area.

Equations (2) to (4) provide an expression for the radiated sound power:

$$P = \text{Re}(Z_2) \left(\frac{F_{1,rms} H}{|Z_1|} \right)^2, \quad (5)$$

which does not contain the velocities of the vibrations in the structure. From this equation we can draw up a

'recipe' for reducing the noise output from an electrical appliance:

- reduce the excitation force F_1 ,
- increase the mechanical input impedance Z_1 of the structure,
- prevent the propagation of the vibration by decreasing the transfer function H , and
- reduce the acoustic radiation by decreasing the real part of the mechanical radiation impedance Z_2 .

The first three items produce the most effect, since the associated factors occur as the square in equation (5).

At Philips Research Laboratories acoustic investigations have been carried out on various Philips products, including vacuum cleaners, food processors, fluorescent lighting switchgear, Compact Disc players and television deflection coils. The theoretical considerations outlined above will be illustrated by a description of our investigation of the noise generated by a washing machine and a refrigerator.

The washing machine whose noise output was investigated was a front-loader, i.e. one in which the wash is contained in a drum accessible from the front. The drum rotates in a 'tub', which can be partly filled with water or suds; see *fig. 1*. In a complete wash programme the machine works through the three main modes of operation: wash, at a drum speed of 50 revolutions per minute; rinse, at the same speed; spin drying, at a maximum speed of 850 revolutions per

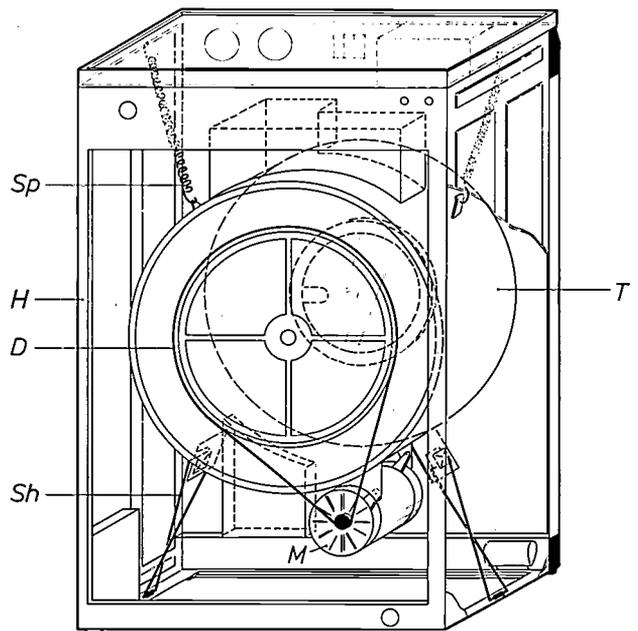


Fig. 1. The interior of a front-loading Philips washing machine. *Sp* spring. *H* cabinet. *D* V-belt drive. *Sh* shock-absorber. *M* motor. *T* tub, filled with water or suds. The drum (not shown) rotates in the tub. The view is from the back of the machine. The wash is put into the tub from the other side of the machine.

minute. (In the latest Philips washing machines this speed can be even higher.)

In a washing machine several sources of noise can be identified, such as:

- the motor,
- the V-belt drive and the ball-bearings of the drum,
- the wash moving in the liquid, and
- the pump for draining the suds or rinsing water.

This article will be mainly concerned with the motor and its power supply. We shall see that the thyristor speed-control system is a major source of noise. Power-supply modifications and some changes in design have made the latest Philips washing machines, such as the AWB 991, very much quieter, see *fig. 2*.

The main source of noise in a refrigerator is the compressor unit; see *fig. 3*. The compressor increases the pressure of the gaseous refrigerant, a fluorocarbon called Freon, from 1.2 bars to about 14 bars. The compressed gas is liquefied in a condenser outside the actual refrigerating compartment. The pressure of the liquid refrigerant decreases in a capillary restrictor, and the liquid is then returned to the gaseous form in the evaporator against the rear wall of the refrigerating compartment. This is the process for extracting heat from the refrigerating compartment.

In our investigation we treated the compressor unit as an independent device, containing various separate

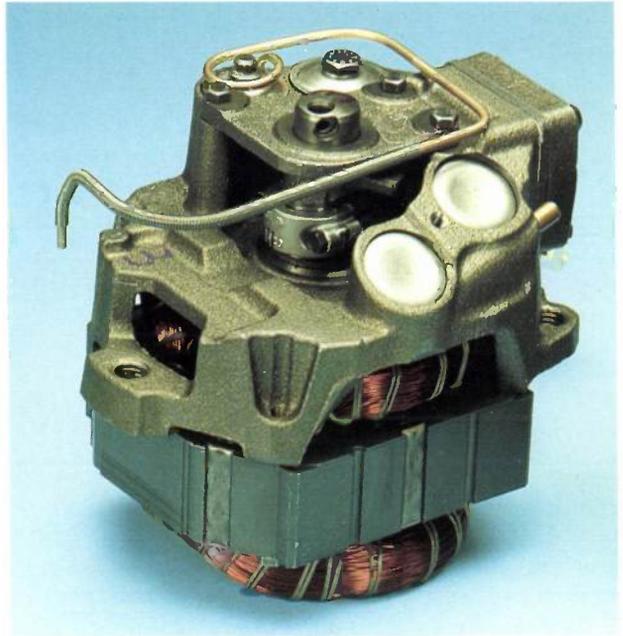


Fig. 3. The interior of an earlier version of one of the Philips compressor units for refrigerators. The two circular expansion chambers of the intake silencer can be seen at the top right. Units of this type, in an improved version, are produced in large numbers by the Philips factories in Cassinetta, Italy.



Fig. 2. The Philips AWB 991 washing machine. In this machine the speed of the drum is 1200 revolutions per minute during spin-drying.

noise sources. The most important ones are:

- the intermittent induction intake of the gas,
- the alternating forces on the piston due to compression of the gas, and
- the intermittent discharge of the gas under pressure.

In this article we shall only consider the intake of the gas, one of the main sources of noise in refrigerator compressors until recently. It will be shown that the noise from this source has been considerably reduced in intensity by modifying the intake silencer (silencers are sometimes called 'mufflers'). The proposed modification has given a reduction of 3 dB(A)^[3] in the noise level from the compressor unit built into many Philips refrigerators, such as the ARG 197 refrigerator-freezer; see *fig. 4*.

The intake silencer was improved by optimizing its operation as a reactive silencer. A silencer of this kind usually consists of a number of expansion chambers as in *fig. 3*, linked by tubes. The noise is attenuated in the silencer because the acoustic input impedance of the silencer is mainly reactive and not matched to the internal impedance of the source. Because of this mismatch, the sound source delivers less power than without the reactive silencer. The action of the reactive si-

[2] The 'r.m.s. value' is understood here to be the root-mean-square value in a narrow frequency band. Defined in this way the r.m.s. value is in general frequency-dependent.

[3] The 'A' in dB(A) indicates that a weighting is applied at each frequency from a standardized curve, the A-weighting curve, which correlates with the sensitivity of the human ear.



Fig. 4. The Philips ARG 197 refrigerator-freezer. The refrigerator is at the top, the freezer below. The appliance has only one compressor unit and has electronic temperature control, with a temperature sensor in both the refrigerator and freezer compartments. The control circuit operates a valve that divides the refrigerant between the two evaporators in the two compartments.

lencer can also be considered as the result of destructive interference of incident and reflected sound waves. The waves are reflected from one of the successive discontinuities in the silencer. The principle of the reactive silencer has also been used to suppress the exhaust noise from Philips vacuum cleaners^[4].

The analysis of the vibration in the washing machine will be considered first, and then the measures taken to reduce the vibration. The same procedure will be followed for the refrigerator compressor.

The washing machine

General

Before making any acoustic measurements or calculations it is first advisable to listen carefully. When we listen to a washing machine a number of clearly identifiable sounds can be heard. These originate from sources such as the pumps and the wash and spin-drying operations. Any assessment of these sounds has to include a rather subjective view of the amount of annoyance and their duration. It turns out that washing makes less noise than spin-drying, but

that the noise is more annoying and lasts longer. The noise from the pumping is also annoying, but does not last very long. This can be reduced fairly easily by mounting the pump on vibration isolators.

In view of the foregoing the discussion here will be limited to our investigation of the noise produced during the wash operation. In the washing machine we investigated the drum is driven by a d.c. motor that has a speed of 800 rev/min during the wash cycle. The drum speed is then 50 rev/min. The stator field in the motor is produced by two permanent magnets. The motor was designed for use in washing machines^[5]; the torque and hence the speed of the motor can easily be varied by changing the mean value of the current in the rotor windings. The same motor can therefore be used for wash and spin dry.

The mean current in the rotor windings is varied by means of an electronic circuit containing thyristors. This circuit rectifies the alternating current from the mains and prevents the flow of current during a preset part of the period of 10 ms; see *fig. 5a*. The motor speed is measured by means of the e.m.f. induced in the rotor windings. A control circuit compares the actual speed with the desired value. Any difference results in a change of the time at which the thyristors are triggered. The circuit thus gives a sudden increase in the rotor current once every 10 ms.

These periodically repeated sharp increases in the rotor current and the corresponding discontinuities in the motor torque could be a major cause of the noise produced during the wash cycle. To verify this hypothesis we measured the noise spectrum during the wash operation under two conditions: one with the normal supply from the mains via thyristors, and the other with the motor receiving its power from a stabilized direct-current source.

The measurements were performed in the special non-rectangular reverberation chamber at Philips Research Laboratories in Eindhoven^[6]. The large number of reflections from non-parallel walls in this chamber give a virtually diffuse sound field, with a simple relation between the sound pressure and the power radiated by a source. The r.m.s. sound pressure at various frequencies was measured with 'one-third octave band filters'. The measurements were converted into sound power and plotted in dB. In addition to each spectrum, the noise power summed over all the frequency bands measured was also plotted in dB(A).

The noise spectra measured in this way are shown in *fig. 5b* and *c*. Comparison shows that the noise reduction due to stabilizing the power supply is very considerable: the total noise power has fallen by 13 dB(A). The most important source of noise during

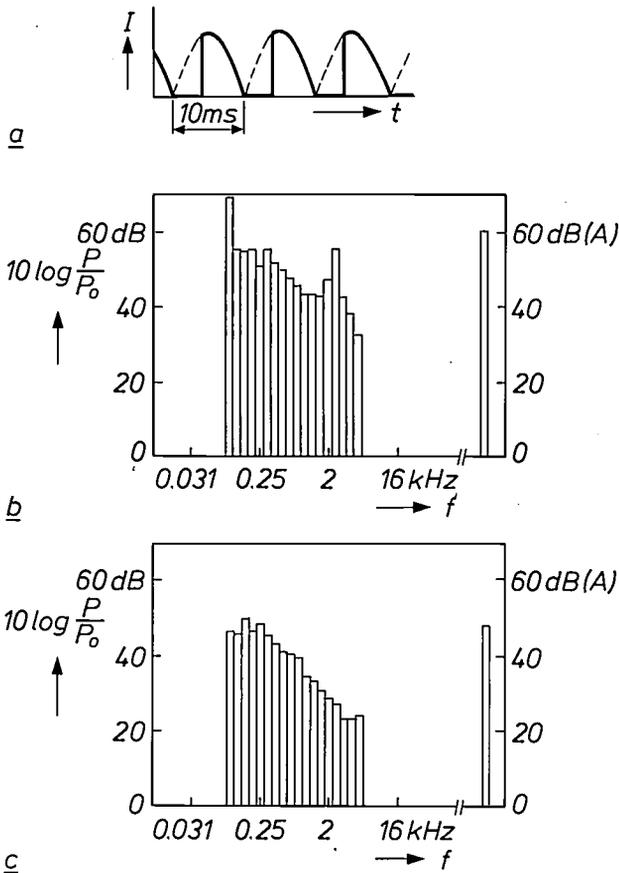


Fig. 5. *a*) The motor current I as a function of time t . During a part of a period of 10 ms (half the period of the alternating current from the mains) the flow of current is prevented by a thyristor; it then increases suddenly. *b*) Noise spectrum of the washing machine in fig. 1, measured during the wash cycle. The spectrum shows the ratio in dB of the noise power level P to a reference level P_0 of 10^{-12} W as a function of frequency f . Each bar indicates the noise power measured within a band of $1/3$ octave. The shaded bar relates to the $1/3$ octave band at 2.5 kHz. The bar on the far right relates to the ratio to P_0 in dB(A)^[3] of the noise power summed over all frequencies. *c*) A noise spectrum measured in the same way but with a stabilized direct voltage for the motor.

the wash cycle has thus been identified: vibration in the motor due to the sudden rise in the motor current every 10 ms.

Analysis of the motor vibration

After the main source of the noise produced during the wash cycle has been identified, it is necessary to gain a better understanding of the mechanism of the vibration in this source. Only then will it be possible to take the appropriate measures to reduce this vibration. We started by measuring the noise spectrum of the motor alone, running at the speed used during the wash cycle.

Fig. 6*a* shows a cross-section of the motor. The measured noise spectrum is shown in fig. 6*b*. If we compare the spectrum with that of the washing machine in operation (fig. 5*b*), we see that both spectra have a peak (shaded) in the $1/3$ -octave band at 2.5 kHz.

Our next step was to attach two accelerometers to the housing H_0 of the motor, as indicated in fig. 6*c*. We struck the housing with a hammer and recorded the two damped oscillations; see fig. 6*d*. The resonant frequency of the stator is the same for both signals: 2600 Hz. This value corresponds to the peak in the noise spectrum in the 2.5-kHz $1/3$ -octave band. The measured acceleration amplitudes and phase difference suggest that the permanent magnets oscillate in opposition in this resonance. The tubular housing then acts as a relatively weak spring linking the magnets.

Further comparison of figs 5*b* and 6*b* shows that the noise power in the $1/3$ -octave band at 2.5 kHz is 3 dB higher in the spectrum of the washing machine. In the transmission of the motor vibration to the tub and the cabinet something seems to be added to the power. However, the cabinet is in fact intended to provide sound insulation for the sound radiated from the motor. In other frequency bands, especially in the low range, we find even higher contributions to the sound power. It is interesting to consider the part played by the tub, to which the motor is directly attached. Since the tub is fairly large — diameter 50 cm and width 30 cm — it is a good radiator of low-frequency sound.

Analysis of the tub vibration

During the wash cycle we measured the acceleration of the tub, with the accelerometer mounted on its rim; see fig. 7*a*. We see that the vibration spectrum consists mainly of harmonics of 100 Hz. These harmonics are excited by the pulses in the motor current. For a closer analysis we mounted a second accelerometer on the other side of the tub. During the wash we recorded the two signals simultaneously and analysed them with a two-channel FFT analyser (FFT: Fast Fourier Transform). This analyser can measure both the amplitudes and the phase difference of two signals in a particular frequency range.

The FFT analysis shows that in the range below 600 Hz the 100-Hz harmonics of both signals are equal in amplitude but opposite in phase. This suggests that the vibration of the tub is a 'rigid-body oscillation'; see fig. 7*b*. At low frequencies the tub assembly with its sprung suspension can therefore easily be approximated by a mass-spring system. Significant resonances from the walls of the tub only appear in the range from 600 to 900 Hz. Fig. 7*c* shows for

[4] Plastics in vacuum cleaners, Philips Tech. Rev. 44, 43, 1988.

[5] R. Raes and J. Schellekens, A speed-controlled d.c. motor for a washing machine, Philips Tech. Rev. 34, 163-169, 1974.

[6] J. M. van Nieuwland, A. Petterson and C. Weber, The design and construction of a non-rectangular reverberation chamber, Philips Tech. Rev. 37, 176-179, 1977.

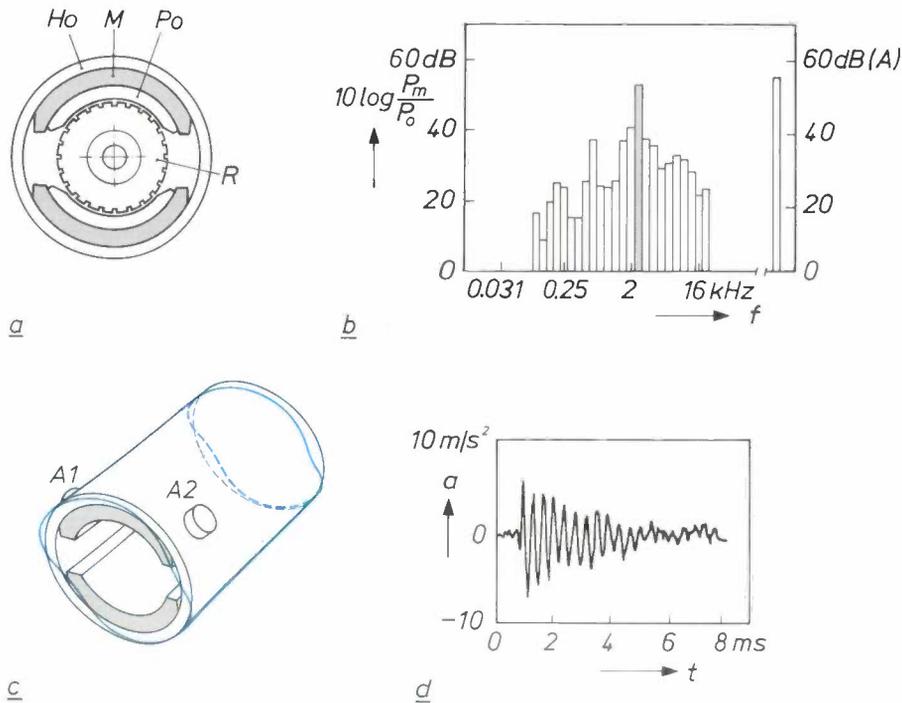


Fig. 6. *a)* The washing-machine motor. *Ho* housing. *M* permanent magnets. *Po* pole pieces. *Ho*, *M* and *Po* form the stator. *R* rotor. *b)* The noise power P_m of the separate motor in dB as a function of frequency f ($P_0 = 10^{-12}$ W). The shaded bar relates to the 1/3-octave band at 2.5 kHz. *c)* The blue lines represent schematically the relative movements of the stator halves during a resonance at a frequency of 2.6 kHz. *A1* and *A2* accelerometers. *d)* The acceleration a measured by one of the transducers as a function of time t .

example the pattern of flexural vibrations at 630 Hz at the circumference of the tub. The asymmetry of the pattern is due to an irregular distribution of the mass and stiffness of the material at the circumference.

Reducing the vibration

One of the most obvious measures here is to modify the steep leading edges in the motor current. We proposed a modification of the power supply to give a double step in each leading edge, as in *fig. 8a*, with the time between the two steps equal to half a period of the stator oscillations as in *fig. 6c*. The result of the double step is that two oscillations with a phase difference of 180° are excited every 10 ms, as shown in *fig. 8b*. With this phase difference these two oscillations should in theory cancel each other.

In practice the two oscillations do indeed largely cancel out, as *figs 8c* and *d* show. These figures give the results of stator-acceleration measurements for a supply with no steps in the current pulse and with a stepped pulse. The maximum acceleration is much smaller with the modified current pulse.

In the latest Philips washing machines a slightly different solution has been adopted for the problem of the motor power supply. The thyristor speed control

has been replaced by a chopper circuit, in which the sinusoidal mains voltage is rectified and then converted by power transistors into pulses at a repetition

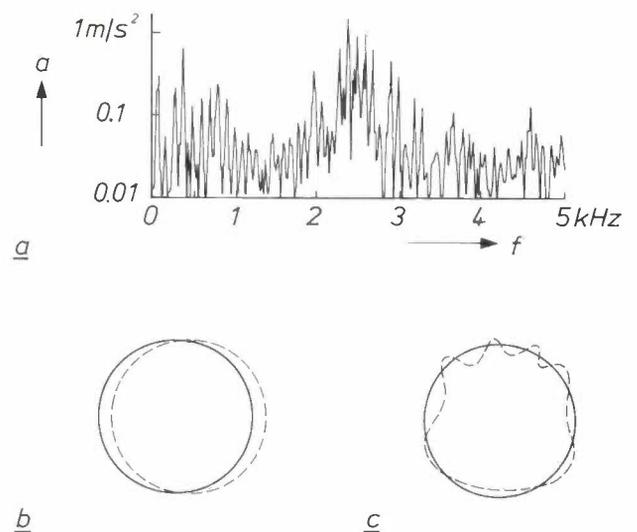


Fig. 7. *a)* The measured r.m.s. value a of the acceleration of the tub (T in *fig. 1*) as a function of frequency f . *b)* Schematic indication of the movement of the tub during 'rigid body' oscillations. *c)* The pattern for flexural vibrations at 630 Hz at the circumference of the tub.

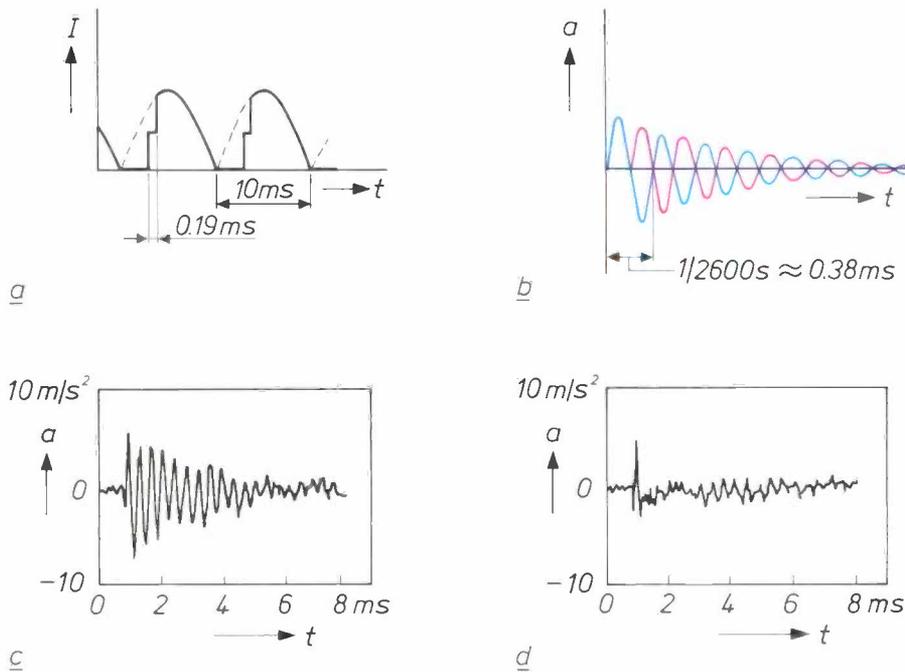


Fig. 8. *a*) The motor current I as a function of time t with a modified power supply for the motor. A double step appears in the leading edge every 10 ms. *b*) The damped oscillations of the stator halves, as would be expected; see fig. 6c. a acceleration. t time. The oscillations shown in red and blue have a phase difference of 180° and are excited by the first and second steps in I . The time of 0.19 ms between the two steps is made approximately equal to half a period of the oscillations at a frequency of 2600 Hz. *c*) The measured oscillations of the stator with the unmodified power supply for the motor, and *d*) with the modified power supply.

rate of about 10 kHz. The width of the pulses is variable. Because of the filtering action of the inductive motor windings the current is virtually a direct current. The mean value of the current and hence the speed of the motor are controlled by varying the width of the 10-kHz pulses. In this circuit the current therefore changes much less abruptly than it did with the thyristor-controlled supply.

Another approach was to improve the motor mounting. Simple rubber rings were replaced by more effective vibration isolators. This and the power-supply modification described above had the effect of substantially suppressing both the actual motor vibration — 100-Hz harmonics and 2600-Hz resonance — and its propagation.

We also improved the sound insulation of the cabinet. In the transmission of the sound from the interior of the machine to the outside world via the casing, three transmission paths are important:

- transmission via the layer of air between tub and cabinet,
- transmission via the springs and shock-absorbers that carry the tub, and
- airborne transmission via holes in the sheet steel of the cabinet.

In the first path the air layer between the tub and the cabinet acts as a relatively stiff spring for the low-frequency 'rigid-body oscillations' of the tub. This spring transmits the low-frequency oscillations of the tub to the sheet-steel sides of the cabinet with very little change in amplitude and vibration pattern. These sheet-steel 'plates' therefore go into a forced flexural vibration, so that large parts of the 'plates' oscillate in phase. The mechanical radiation impedance for flexural vibrations of this kind is high, so that the sound insulation of the cabinet in this situation is almost zero. As we have seen above, the magnitude of the low-frequency oscillations has fortunately been considerably reduced by the measures proposed.

In sound transmission by the second path the sides of the cabinet are excited at the places where the springs and shock-absorbers are fixed. These local excitations set up free flexural vibrations, which differ from the forced type in that different parts of a 'plate' oscillate in antiphase. The wavelength of free flexural low-frequency vibrations is much smaller than that of the airborne sound, so that pressure differences close to the 'plate' are equalized. The acoustic radiation impedance for these free flexural vibrations is thus low. The low-frequency oscillations of the tub shown in

fig. 7b and c will therefore be almost unable to reach the ambient air by this path.

If the wavelength of the free flexural vibrations in the 'plate' is smaller than that of the radiated sound, the sound waves near the plate will be 'acoustically short-circuited'. The air layers furthest removed from the vibrating plate are then no longer disturbed: the air vibrations die out. The wavelength of sound is inversely proportional to the frequency; the wavelength of free flexural vibrations is inversely proportional to the root of the frequency. The acoustic-short-circuiting effect thus decreases with rising frequency, and vanishes when the wavelength of the air vibrations becomes equal to the wavelength of the free flexural vibrations. The corresponding frequency is called the 'critical frequency' [7]. It is 15 kHz for 1-mm sheet steel, the material generally used for washing-machine cabinets.

In the third path the sound is transmitted to the ambient air through openings in the cabinet. It therefore makes sense to 'plug the leaks' in the cabinet as far as possible. A significant leak is the gap between the cabinet and the floor. Closing this gap is therefore an important noise-reduction expedient. A proposal has also been made to fit absorbent material to dissipate acoustic energy inside the cabinet.

The last, but by no means the least-significant, measure we have proposed relates to a modification of the motor end-shields. These are mounted at the ends of the stator and support the rotor bearings. Openings in the end-shields admit the air for cooling the rotor windings. It has been found that the rotor does not in fact become too hot if these openings are closed. The noise output from the motor is then much lower, however. The effect of running with these openings closed is particularly valuable in the spin-dry cycle, when the motor speed is higher.

Applying the measures mentioned reduces the noise output during the wash by about 5 dB(A).

The refrigerator compressor

General

Fig. 9 shows a cross-section of the compressor unit, which consists of the compressor proper, an electric motor and a hermetically sealed housing of 2.5-mm steel. This housing is called the 'shell'. The reciprocating motion of the piston in the compressor is obtained from crank gear. At every revolution of the crankshaft the intake and outlet valves are opened alternately by the reduced pressure and the excess pressure of the gas in the cylinder.

The motor shaft is hollow, and its lower end dips into a quantity of oil in the bottom of the shell. The hole in the shaft widens, so that oil is drawn into the shaft by centrifugal force. The oil is forced into the crankshaft and from there through the hollow connecting rod to the gudgeon pin. The oil thus lubricates the crankshaft bearings, the two connecting-rod bearings and the piston in the cylinder. Oil is also sprayed from the crankshaft around the cylinder to cool it.

The compressor is spring-mounted in the shell. The Freon gas enters the free space between the shell and the compressor through an inlet channel. The gas is drawn into the cylinder from this space through an

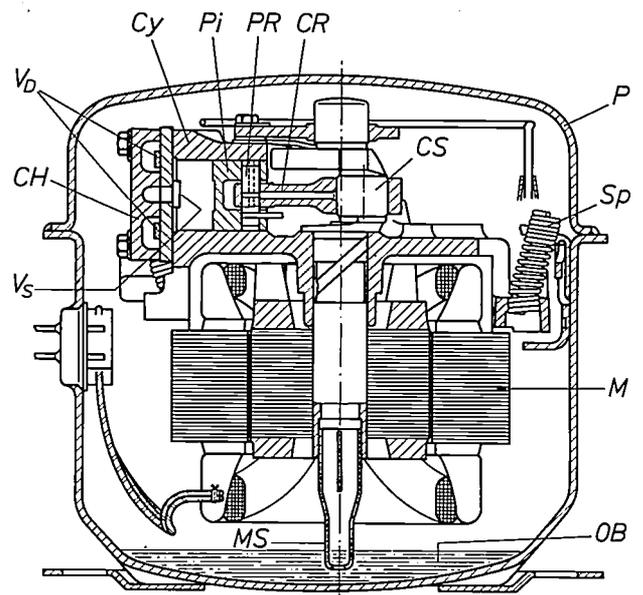


Fig. 9. The compressor unit. V_s inlet valve. CH cylinder head. V_b outlet valve. Cy cylinder. Pi piston. PR gudgeon pin. CR connecting rod. CS crankshaft. P shell. Sp spring. M motor. OB oil bath. MS hollow motor shaft with 'oil pump': the opening widens towards the top.

intake silencer with two expansion chambers in cascade; see fig. 3. The compressed gas leaves the compressor unit through an output silencer and a metal tube, the discharge tube. This is made as flexible as possible and passes through a gas-tight seal in the wall of the shell.

The noise sources in the compressor unit, which were mentioned in the introduction, can transfer sound to the wall of the shell via

- the Freon gas,
- the oil reservoir,
- the discharge tube, and
- the three springs of the suspension system.

Finally, the shell wall radiates the sound to the surrounding air.

In our investigation we examined the various possible sources of noise. This entailed first measuring the noise produced both with the source and without the source, if possible. This often meant that modifications had to be introduced. For example, we were able to study the effect of parts of the crank gear by operating the compressor with a lighter piston and with a stiffer connecting rod.

We compared the results of the measurements with the results of calculations based on models. When the results of the calculations were in reasonable agreement with the measurements, the models helped us to find ways of reducing the noise levels.

The investigation of the various sources of noise led eventually to seven proposed modifications to reduce the noise levels. These gave a total reduction of 10 dB(A), without any need to change the functional specifications of the compressor. The design of the model for the noise due to the gas intake will now be considered, as well as the measurements made in investigating this noise.

The model for the intake noise

The intermittent volume flow resulting from the opening and closing of the inlet valve must be considered as the actual source of the intake noise. This flow causes pressure fluctuations that propagate in the intake silencer in the direction of the intake opening. The pressure fluctuations excite acoustic resonances in the gas-filled space in the shell.

This hypothesis can be verified by analysing the oscillations of the wall of the shell after a 'cold start'; this is done by recording a number of frequency spectra in succession. The temperature of the gas in the compressor unit rises from 20 to 90 °C after the cold start. The velocity of sound in the gas in the shell then changes such that:

$$\frac{c_1}{c_2} = \frac{\sqrt{T_1}}{\sqrt{T_2}}, \tag{6}$$

where c_1 and c_2 are the sound velocities in the gas at the absolute temperatures T_1 and T_2 . The wavelength $\lambda = c/f$ of a particular resonance in the gas depends only on the geometry of the space. The resonant frequency f therefore changes with temperature in the same way as the sound velocity. Since the principal resonance peaks in the measured spectra shift to a higher frequency when the gas temperature increases, these peaks are the result of acoustic resonances in the gas. These resonances in turn set the wall in motion. They are not therefore resonances of the wall itself.

For the model that describes the origin, transfer and radiation of the intake noise, we can write the equation:

$$P_{\text{rad}} = \frac{P_s}{IL_d \cdot IL_b}, \tag{7}$$

where P_{rad} is the power of the sound radiated by the shell wall, IL_d is the insertion loss for the sound transmission in the intake silencer and IL_b is the insertion loss for the sound transmission via the gas-filled space and the shell wall. P_s is the power of the sound originating at the inlet valve for the case in which the sound can be directly radiated to the environment, i.e. without silencer and shell. All the quantities in eq. (7) (and those in the next equation) are a function of the frequency f .

For P_s we can write the equation:

$$P_s = Q_{\text{rms}}^2 \text{Re}(Z_s), \tag{8}$$

where Q_{rms} is the r.m.s. value of the intermittent volume flow at the inlet valve [2] and Z_s is the acoustic radiation impedance for the sound source formed by the intermittent volume flow in the opening of the inlet valve (without the silencer and shell).

This sound source can be considered as being replaced by a rigid flat piston of radius r , which moves to and fro inside an infinite stationary wall (baffle). In an electrical analogue Z_s can then be approximated by the impedance of a network containing two resistances, a capacitance and an inductance [8]; see fig. 10. The inductance in this network is equivalent to the mass of the gas, the capacitance to its elasticity. The resistances show that energy is in fact transferred to the gas. R_1 , R_2 , C and L can be expressed in terms of ρ , the density of the gas, c and r .

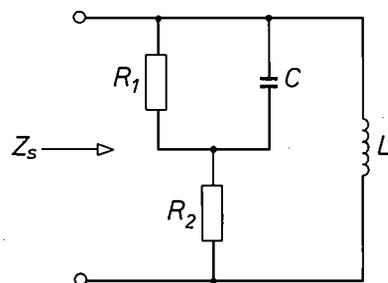


Fig. 10. Electrical analogue for a reciprocating piston in an infinite baffle. The impedance Z_s of the network is analogous to the acoustic radiation impedance of the sound source formed by the moving piston. R_1 and R_2 resistances. C capacitance. L inductance.

[7] See pages 290 to 294 of: L. L. Beranek, Noise reduction, McGraw-Hill, New York 1960.

[8] See pp. 118 ff. of the book by Beranek, note [1].

The sound power at the inlet valve can be calculated as a function of frequency (eq. 8) if the variation of the volume flow as a function of time is known. We used two methods for determining this function $q(t)$. The first is based on the assumptions that the lift of the inlet valve is an exponential function with time constant τ , that the pressure difference Δp across the valve decreases linearly with time after it has been

flow p_i and q_i before the element and p_{i+1} and q_{i+1} after the element:

$$\begin{pmatrix} p_{i+1} \\ q_{i+1} \end{pmatrix} = H_i \begin{pmatrix} p_i \\ q_i \end{pmatrix}. \quad (10)$$

The transfer matrices for the various elements of the silencer can be calculated with the one-dimensional acoustic wave equation. The transfer matrix H of the complete silencer is given by the product of the matrices of n separate elements:

$$H = \prod_{i=1}^n H_i.$$

The insertion loss IL_d can be calculated from the total transfer matrix H and the acoustic radiation impedance. A computer program was written for these calculations, but this will not be discussed here. It can be

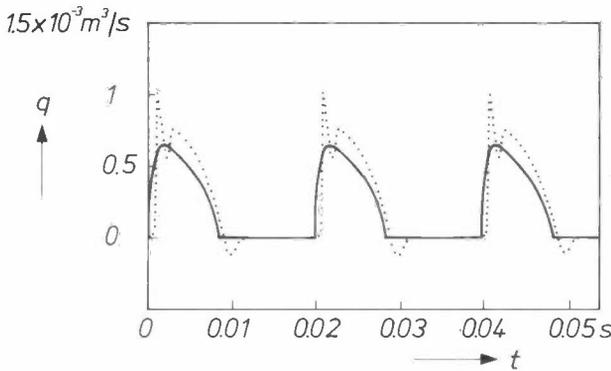


Fig. 11. Two different results of calculations of the volume flow q at the inlet valve of the compressor as a function of time t . The continuous line relates to the result of the calculation from eq. (9), which is based on a number of simplifications in the behaviour of the valve and the gas. The dotted line relates to the result of a computer simulation.

opened, and that the volume flow is proportional to $\sqrt{|\Delta p|}$. From these assumptions it follows that:

$$q(t) = \frac{A}{\rho} (1 - e^{-\frac{t}{\tau}}) \sqrt{t_0 - t} \text{ for } 0 \leq t \leq t_0, \quad (9)$$

where A is a constant that depends on the capacity of the compressor (2.9 kg per hour) and t_0 is the time the inlet valve is open. The quantities τ and t_0 can be estimated from measurements of the height of the valve as a function of time.

In the other method of determining $q(t)$ a computer program simulates the thermodynamic behaviour of the compressor^[9]. Fig. 11 shows the results of both methods of calculation. The computer simulation probably gives a more realistic picture, since it takes the dynamic properties of the inlet valve into account. These can cause fluctuations in the gas flow, which may even become negative.

The insertion loss IL_d of the intake silencer is equal to the ratio of the intake noise power without the silencer to the intake noise power with it. IL_d can be calculated by setting up a transfer matrix H_i for each element of the silencer, which can be an expansion chamber or a connecting tube. This matrix gives the relation between the sound pressure and the volume

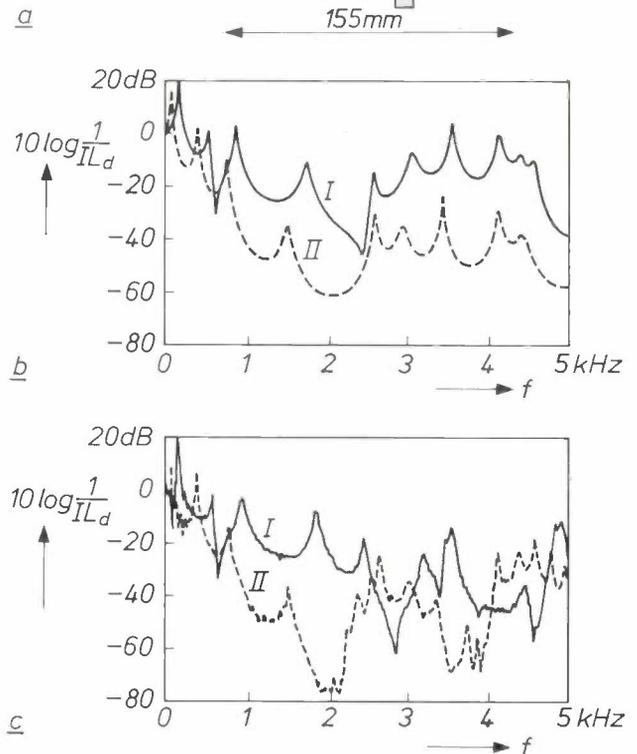
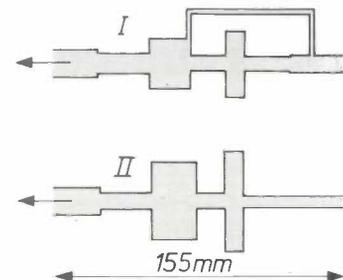


Fig. 12. The results of calculations and measurements for two different silencers. a) Schematic representation of an earlier version (I) and a recent version (II) of the silencer. b) the reciprocal of the calculated insertion loss IL_d , and c) the reciprocal of the measured insertion loss IL_d , both in dB as a function of frequency f for silencers I and II.

used for comparing the insertion losses of different versions of the intake silencer.

To calculate the insertion loss IL_b for the shell and the free gas volume models would have to be set up for the acoustic resonance of the gas volume, the mechanical resonances in the steel wall of the shell, the coupling between the two kinds of resonance, and the sound radiation from the shell. This could be done by using the finite-element method. However, in this case such calculations would be complex and time-consuming, so we decided to do without and just make a number of measurements.

The calculations and measurements

Fig. 12 shows the results of calculations and measurements of the insertion loss IL_d of two intake silencers. Silencer *I* is an earlier version (see fig. 3) and silencer *II* is the most recent one, introduced largely because of the investigations described in this article. The results of the calculations given in fig. 12b are in good agreement with the results of the measurements given in fig. 12c. Silencer *II*, which has larger expansion chambers and no parallel tube, has a better insertion loss at nearly all frequencies. At very low frequencies, incidentally, this quantity is found to be smaller than 1, so that the sound is amplified. Other measurements showed that the performance of silencer *I* is even worse if the parallel tube is omitted.

The most obvious method of measuring the insertion loss IL_b for the shell and the free gas volume is to place a small sound source at the mouth of the intake silencer, and then to measure the sound with and without the shell. The problem here is that it is difficult to find a small source that will produce enough sound to give a sufficiently large signal-to-noise ratio outside the shell. We therefore made use of the reciprocity principle, which in this case states that the insertion loss can also be measured by interchanging the source and the measuring device. For our measurements in the reverberation chamber we used two large loudspeaker units and measured the sound at the mouth of the intake silencer with a small microphone. The insertion loss IL_b as a function of frequency is the ratio of the power spectrum measured without the shell to the spectrum measured with it (including the Freon gas).

Conclusion

The previous section showed in general terms how the power of the sound produced by the intermittent volume flow at the inlet valve could be calculated. The calculation of the insertion loss of the intake silencer has also been treated. The insertion loss for the shell and the free gas volume was measured in the way just

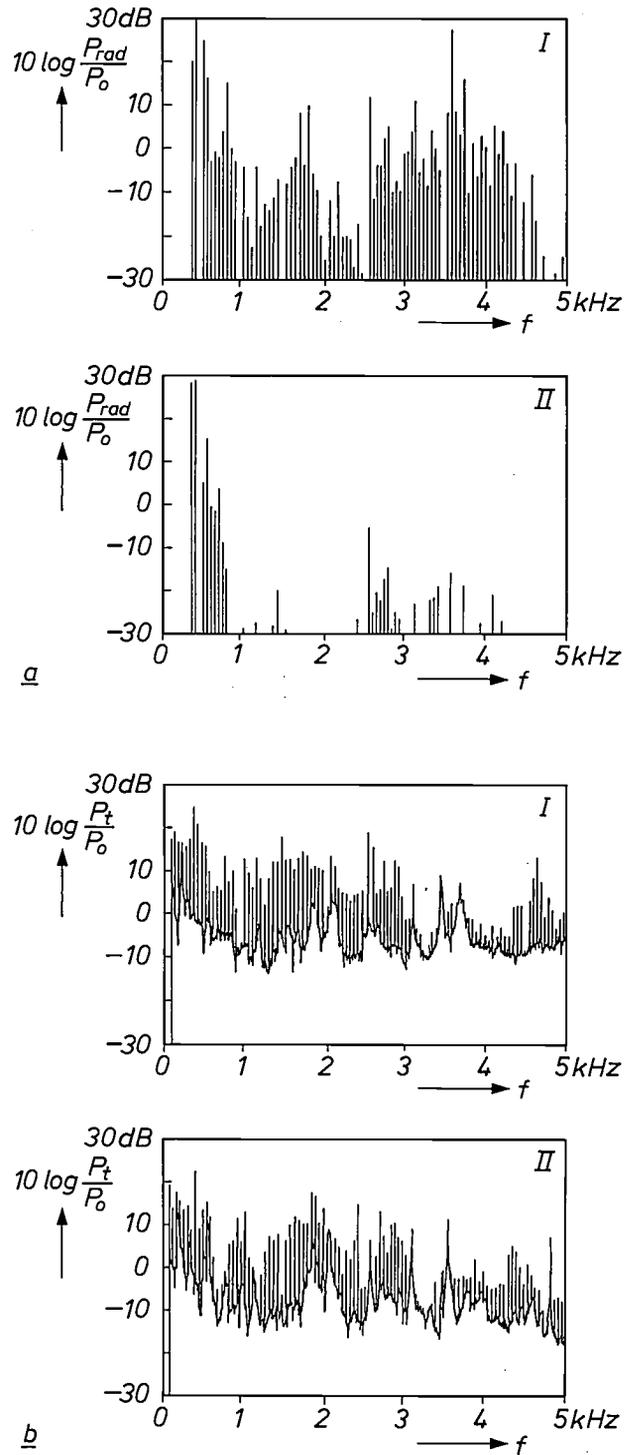


Fig. 13. *a*) Calculated power P_{rad} of the intake noise in dB ($P_0 = 10^{-12}$ W) as a function of frequency f for silencers *I* and *II*. *b*) The total measured noise power P_t of the compressor in dB, also for silencers *I* and *II*.

[9] L. J. M. Kuypers, G. A. A. Asselman and G. van den Berk-mortel, Reed valve simulation: a comparison of numerical and experimental results, Proc. 15th Int. Congr. on Refrigeration, part 2, Venice 1979, pp. 795-801.

described. The three quantities that determine the power P_{rad} of the intake noise as a function of frequency f in equation (7) are thus available. $P_{\text{rad}}(f)$ is found by multiplying the three spectra for IL_d^{-1} , IL_b^{-1} and P_s . A program that also calculates a Fourier transform of the volume flow $q(t)$ as given by eq. (9) performs this multiplication fairly easily.

Fig. 13a shows the results of calculations for the compressor with intake silencer *I* and for the compressor with intake silencer *II*. Here again, the considerable improvement achieved by modifying the silencer can be seen. Fig. 13b shows the measured power spectra of the total noise produced by the compressor unit. The calculated and measured spectra for silencer *I* in particular demonstrate that the intake noise is an important part of the total noise output. The two spectra for silencer *II* differ more than the spectra for silencer *I* because the other noise sources become more significant when the silencer has been improved.

It also appears from the agreement between the calculated and measured spectra for silencer *I* that the

models give a fairly good approximation of the behaviour of practical devices. The models can therefore be used for further improvements in the design of the intake silencer, and for reliable predictions in the future design of other kinds of reactive silencer.

Summary. Acoustic noise generated by a source inside an electrical appliance reaches the environment after the successive steps of excitation, transmission and radiation. Noise reduction can be achieved by tackling each of these steps separately. In a washing machine the motor is one of the main sources of noise. Measurements show that this noise originates mainly in the periodic abrupt change in the motor current from the thyristor-controlled power supply. A different power-supply circuit, improved motor supports and a sound-proofed cabinet have reduced the noise from the washing machine by 5 dB(A) during the wash cycle. In the compressor unit of a refrigerator the main noise source is the intermittent intake of the Freon gas through the inlet valve. Calculations based on models have given a better understanding of the origin and transmission of this noise. Measurements show that the models correspond well to the behaviour of practical devices. Modifying the reactive intake silencer of the compressor unit gives a noise reduction of 3 dB(A).

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale, P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

Ned. Philips Bedrijven B.V.
PHILIPS NATUURKUNDIG LAB
BIBLIOTHEEK WY - 1
P.O. Box 80.000
5600 JA EINDHOVEN
NEDERLAND

K. Murata (<i>Univ. Osaka</i>) & D. F. Kyser	D. F. S	Monte Carlo methods and microlithography simulation for electron and X-ray beams	Adv. Electron. & Electron Phys. 69	175-259	1987
F. Weling & H. Wehr	A	Properties of multiple clad fibers: experiments and numerical predictors	Appl. Opt. 26	4852-4856	1987
B. A. Joyce, J. Zhang (<i>Imperial College of Sci. & Technol., South Kensington</i>), J. H. Neave & P. J. Dobson	R	The application of RHEED intensity effects to interrupted growth and interface formation during MBE growth of GaAs/(Al,Ga)As structures	Appl. Phys. A 45	255-260	1988
D. E. Lacklison, G. Duggan, J. J. Harris, C. T. B. Foxon, D. Hilton, C. Roberts & C. M. Hellon	R	Slope resistance characteristics of GaAs-(Al,Ga)As-GaAs single barrier structures	Appl. Phys. Lett. 52	305-307	1988
R. Pandya & A. Martinez	N	Large-area defect-free silicon-on-insulator films by zone-melt recrystallization	<i>ibid.</i>	901-903	1988
K. Shahzad, D. J. Olego & D. A. Cammack	N	Thickness dependence of strains in strained-layer superlattices	<i>ibid.</i>	1416-1418	1988
G. Frens & J. J. F. G. Heuts	E	The double layer potential ϕ_d as a rate determining factor in the coagulation of electrocratic colloids	Colloids & Surf. 30	295-305	1988
P. C. M. Gubbens*, A. M. van der Kraan* (<i>*Interuniv. Reactor Inst., Delft</i>) & K. H. J. Buschow	E	RFe ₁₀ V ₂ compounds studied by ⁵⁷ Fe, ¹⁶¹ Dy, ¹⁶⁶ Er and ¹⁶⁹ Tm Mössbauer spectroscopy	Hyperfine Interactions 40	389-392	1988
B. Gabillard, T. Ducourant, C. Rocher, M. Prost & J. Maluenda	L	A 200-mW GaAs 1K SRAM with 2-ns cycle time	IEEE J. SC-22	693-698	1987
O. Boser & V. Newsome	N	High frequency behavior of ceramic multilayer capacitors	IEEE Trans. CHMT-10	437-439	1987
W. J. van Gils	E	Two-dimensional dot codes for product identification	IEEE Trans. IT-33	620-631	1986
E. Persoon	E	A pipelined image analysis system using custom integrated circuits	IEEE Trans. PAMI-10	110-116	1988
G. Lütteke (<i>Philips Apparatefabrik, Krefeld</i>) & H. C. Raets	A	220-V mains 500-kHz class-E converter using a BIMOS	IEEE Trans. PE-2	186-193	1987
P. Ruterana, J. P. Chevalier (<i>C.E.C.M.-C.N.R.S., Vitry</i>), P. Friedel & N. Bonnet (<i>Univ. Paris-Sud</i>)	L	Cleaning and nitridation of (100) GaAs surfaces: a high resolution electron microscopy study	Inst. Phys. Conf. Ser. No. 87	27-32	1987
C. Schiller, M. Duseaux & J. P. Farges	L	High resolution X-ray diffraction on GaAs and InP substrates	<i>ibid.</i>	621-626	1987

C. Dierieck	B	Pressure potential formulation of 2-D Stokes problems in multiply-connected domains	Int. J. Numer. Methods in Fluids 7	69-85	1987
R. Waser	A	Solubility of hydrogen defects in doped and undoped BaTiO ₃	J. Am. Ceram. Soc. 71	58-63	1988
P. Suchet, M. Duseaux, G. Gillardin, J. Le Bris & G. M. Martin	L	Evidence of the origin of infrared scattering in GaAs with high-resolution infrared tomography	J. Appl. Phys. 62	3700-3703	1987
P. F. Fewster & C. J. Curling	R	Composition and lattice-mismatch measurement of thin semiconductor layers by X-ray diffraction	<i>ibid.</i>	4154-4158	1987
J. Petruzzello, B. L. Greenberg, D. A. Cammack & R. Dalby	N	Structural properties of the ZnSe/GaAs system grown by molecular-beam epitaxy	J. Appl. Phys. 63	2299-2303	1988
R. N. Bhargava	N	Electronic states and pair emission in ZnSe	J. Lumin. 40 & 41	24-27	1988
B. A. Joyce	R	Molecular beam epitaxy: recent trends and future developments	J. Phys. & Chem. Solids 49	237-242	1988
A. Raymond*, R. Piotrkowski*, J. L. Robert*, S. Azema* (*Univ. Montpellier), M. Kubisa*, W. Zadwadzki*, L. Litwin-Staszewska* (*Univ. Warsaw) & J. P. André	L	Impurities in GaAlAs-GaAs heterojunctions metastable character of Si states and magnetodons	J. Phys. Colloq. 48, Suppl. No. 11	C5/243-C5/246	1987
E. K. Broadbent	S	Nucleation and growth of chemically vapor deposited tungsten on various substrate materials: a review	J. Vac. Sci. & Technol. B 5	1661-1666	1987
J. M. Towner	S	The importance of the short-circuit failure mode in aluminium electromigration	<i>ibid.</i>	1696-1700	1987
J.-P. Charlier & P. Van Dooren	B	On Kogbetliantz's SVD algorithm in the presence of clusters	Linear Algebra & Appl. 95	135-160	1987
M. Iost, S. Gourrier, B. Bru, P. Rabinzohn & F. Pasqualini	L	Linewidth control in deep UV contact lithography	Microelectron. Eng. 6	69-75	1987
M. Lemonier & C. Piaget	L	Solid state image sensors for electronic readout of image tubes	Onde Electr. 67, No. 6 (November)	99-107	1987
J. A. Clarke	R	Current trends in optics for projection TV	Opt. Eng. 27	16-22	1988
P. A. Devijver & M. M. Dekesel	B	Learning the parameters of a hidden Markov random field image model: a simple example	Pattern Recognition Theory and Applications, P. A. Devijver & J. Kittler (eds), NATO ASI Series, Vol. F30, Springer, Berlin Heidelberg	141-163	1987
J. Ayers & J. Ladell	N	Spectral widths of the Cu K α lines	Phys. Rev. A 37	2404-2407	1988
D. J. Olego, K. Shahzad*, J. Petruzzello & D. Cammack	N	Depth profiling of elastic strains in lattice-mismatched semiconductor heterostructures and strained-layer superlattices	Phys. Rev. B 36	7674-7677	1987
M. A. Fisher*, A. R. Adams*, E. P. O'Reilly* (*Univ. Surrey) & J. J. Harris	R	Resonant electron scattering due to the central cells of impurities observed in AlGaAs under hydrostatic pressure	Phys. Rev. Lett. 59	2341-2344	1987
F. Berz	R	Effect of barrier thickness on the quantum mechanical transmission across parabolic barriers	Phys. Stat. Sol. b 145	K27-K30	1988
P. A. Devijver	B	Segmentation of binary images using third-order Markov mesh image models	Proc. 8th Int. Conf. on Pattern recognition, Paris 1986	259-261	1986
S. Mukherjee, M. Amato, I. Wacyk & V. Rumennik	N	LD MOS and LIGT's in CMOS technology for power integrated circuits	Proc. IEDM '87, Washington D.C., 1987	778-781	1987
P. Houdy, V. Bodart, C. Hily, P. Ruterana, L. Nevot (<i>Inst. d'Optique, Orsay</i>), M. Arbaoui* N. Alehyane* & R. Barchewitz* (*Univ. Pierre et Marie Curie, Paris)	L	Performances of C-W multilayers as soft X-rays optics: high reflectivity in the range 2d = 60 to 100 Å	Proc. SPIE 733	389-397	1987
M. Rocchi	L	GaAs digital IC's for computer applications	VLSI '87, C. H. Séquin (ed.), Elsevier Science, Amsterdam	221-234	1988



H. J. de Wit and K. Jager, Magnetic domains in amorphous alloys for tape-recorder heads,
PhilipsTech.Rev.44, No.4, 101-109, Sept. 1988.

Certain amorphous cobalt-containing alloys with high saturation polarizations (about 1 T) show promise for applications in tape-recorder heads for video recording at high information densities. The relative permeability μ_r , which has to be very high ($\mu_r > 1000$) for reproducing the stored information, is largely determined by the structure of the magnetic domains. This can be controlled by annealing the alloys in a magnetic field, thus inducing a magnetic anisotropy. The effect on the permeability and the other magnetic properties depends on the direction of the magnetic field and the temperature during annealing, on the form of the material — spun ribbon or sputtered thin film — and on the precise composition.

J. Crucq, Theory and practice of acoustic noise control in electrical appliances,
PhilipsTech.Rev.44, No.4, 123-134, Sept. 1988.

Acoustic noise generated by a source inside an electrical appliance reaches the environment after the successive steps of excitation, transmission and radiation. Noise reduction can be achieved by tackling each of these steps separately. In a washing machine the motor is one of the main sources of noise. Measurements show that this noise originates mainly in the periodic abrupt change in the motor current from the thyristor-controlled power supply. A different power-supply circuit, improved motor supports and a sound-proofed cabinet have reduced the noise from the washing machine by 5 dB(A) during the wash cycle. In the compressor unit of a refrigerator the main noise source is the intermittent intake of the Freon gas through the inlet valve. Calculations based on models have given a better understanding of the origin and transmission of this noise. Measurements show that the models correspond well to the behaviour of practical devices. Modifying the reactive intake silencer of the compressor unit gives a noise reduction of 3 dB(A).

E. W. Meijer, Polyepoxides; formation and properties of their network structure,
PhilipsTech.Rev.44, No.4, 110-121, Sept. 1988.

The polymerization of epoxide monomers with catalysts or initiators results in polyepoxides. Special attention is given to chain polymerization from bis-epoxides: the three-dimensional networks formed permit a wide variety of applications. Methods of gaining a fundamental understanding of the formation of these networks and their structure are illustrated by two examples. In the first example it is shown how phosphorescence spectroscopy can be used to monitor changes in the molecular mobility of polymer segments, both during the formation of the network and as a function of temperature. In the second example it is shown how nuclear-magnetic-resonance spectroscopy can be used to study the mechanism of network formation with the aid of model compounds. Emphasis is placed here on the part played by the $Al(acac)_3$ /catechol initiator-catalyst system.

I wish to subscribe to

PHILIPS TECHNICAL REVIEW

(date)

(signature)

Please tick the appropriate box

Regular subscription 80 guilders or U.S. \$ 35.00 per volume

The subscription includes postage and will start with Vol. 44, No. 1.
Please pay when you receive our invoice.

Student's subscription 32 guilders or U.S. \$ 14.00 per volume

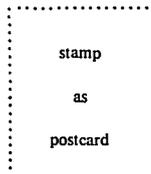
Please send a copy of your student's card or other written proof that you are a student; valid for two volumes.

Name

Initials

Title

Address



**Administration Department
Philips Technical Review**

**Philips Research Laboratories
Building WY 136**

P.O. Box 80 000

**5600 JA Eindhoven
The Netherlands**

Contents

Theory of the GaAs/AlGaAs quantum well	137
R. Eppenga and M. F. H. Schuurmans	
<i>The well-known quantum-mechanics problem of a particle in a box is one of the vital elements in the theoretical description of the quantum-well laser</i>	
Submicron ICs	150
Laboratory-scale manufacture of magnetic heads	151
J. P. M. Verbunt	
<i>The increasing importance of digital signals and the need for higher information densities are having a marked effect on the specifications of magnetic heads and their method of manufacture</i>	
Then and Now (1938-1988)	161
A laser module for 4-Gbit/s optical communications	162
H. Tjassens and J. T. M. Kluitmans	
<i>Simple modifications can extend the useful frequency range of standard dual-in-line packages to well over 4 Gbit/s</i>	
Scientific Publications	165

Theory of the GaAs/AlGaAs quantum well

R. Eppenga and M. F. H. Schuurmans

The advent of the semiconductor transistor in 1948 was a turning point for both applied and fundamental semiconductor physics, and once again we may be on the threshold of a new era today. As dimensions in semiconductor structures become smaller, new effects appear and new theoretical models are devised to describe them. The article below departs from the 'old' semiconductor physics to describe the theory of the quantum well, a layer of GaAs 1 to 10 nm thick between two thicker layers of AlGaAs.

Introduction

Structures consisting of thin layers of semiconductors have special properties that can lead to novel applications. The composition and thickness of the layers and the applied external fields determine the behaviour of the carriers. The number of products that have special features is steadily increasing as more ways are found of modifying the properties of the materials.

Thin crystalline layers (1-10 nm thick) of dissimilar semiconductor materials can nowadays be grown one on the other with a perfect lattice match. Lasers based on thin crystalline semiconductor layers have a lower threshold current than lasers with a thicker active layer, and their emission wavelength is determined by the thickness of the active layer. The mobility of electrons in layered crystalline structures is high, so that transistors can be made for operation at gigahertz frequencies.

A proper understanding of the operation of components consisting of thin crystalline layers requires a detailed quantum-mechanical description of the behaviour of carriers in the material. This behaviour reveals itself in various novel physical effects, such as the quantized Hall effect and the fractional quantized

Hall effect^[1]. Different kinds of layered structures such as quantum wells, superlattices and heterojunctions were dealt with in an earlier special issue of this journal^[2], where their fabrication and general characteristics were described. In this article we shall take a closer look at the theory of the quantum well — a thin crystalline layer of semiconductor material between two thicker layers of another semiconductor material; see *fig. 1*. The theory will focus on the absorption and emission of radiation by a GaAs/AlGaAs quantum well, and is used in practice in model calculations on quantum-well lasers (for example: how is the gain of a

^[1] If a magnetic field is applied at right angles to a current flowing in a conductor or semiconductor, a potential difference can be measured perpendicular to the direction of the current and to the magnetic field. This potential is called the Hall voltage, and the resistance across which it appears is called the Hall resistance. The Hall resistance increases with increasing magnetic field. The Hall effect also occurs in very thin layers of semiconductor material. In this case, however, the Hall resistance does not increase linearly with the magnetic field, but has a number of plateaus. These plateaus occur at resistance values that are an integer number of times e^2/h (this is known as the quantized Hall effect) or at values that are a particular fraction of e^2/h (this is known as the fractional quantized Hall effect), where e is the electronic charge and h is Planck's constant. The occurrence of these levels is connected with the local mobility of electrons under the influence of foreign atoms in the crystal lattice. See for example: B. I. Halperin, The quantized Hall effect, *Sci. Am.* 254, No. 4 (April), 40-48, 1986.

^[2] Special issue on 'Layered semiconductor structures', *Philips Tech. Rev.* 43, 109-165, 1987.

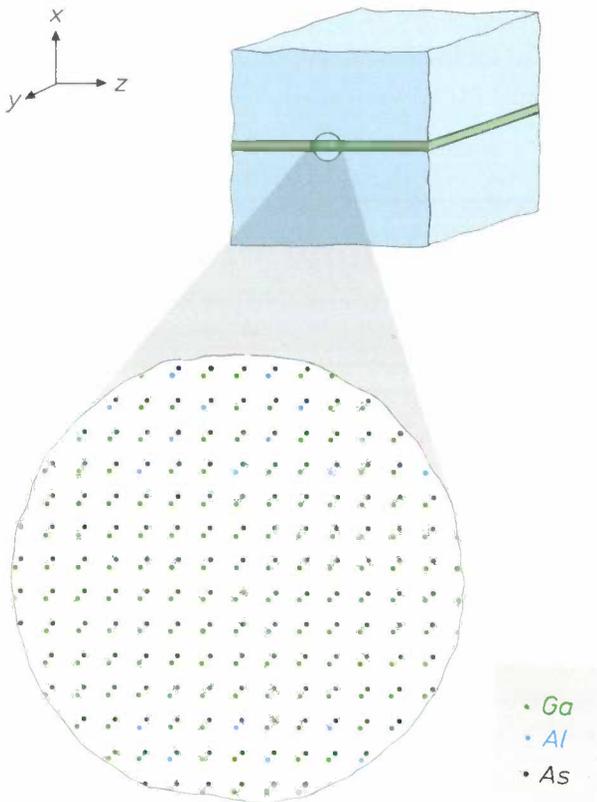


Fig. 1. Schematic representation of a quantum-well structure. A thin layer (1-10 nm) of GaAs (green) is sandwiched between two thicker layers of AlGaAs (blue). Since the distance between the atoms, the lattice constant, is almost identical in GaAs and AlGaAs, these layers join up with a perfect lattice match — with no internal strains or lattice defects.

quantum-well laser affected by the composition of the layered structure? ^[31]). After a general discussion of the behaviour of the carriers in a semiconductor, we shall solve the Schrödinger equation for the quantum well. This gives the energy levels and the states of the electrons in the material. We shall determine these states by means of the 'envelope-function formalism', i.e. we shall only be interested in the overall behaviour of the electrons in the quantum well and not in the behaviour on the scale of the lattice constant. In determining the overall behaviour, however, we do take into account the influence of the atoms at their positions in the lattice. As an example we shall show how the information about states and levels can be used for predicting light-absorption properties.

Semiconductor properties in relation to the quantum well

Before dealing with the physics of the quantum well we shall first give a short general discussion of semiconductor properties. We have to use quantum mechanics in this treatment.

Electrons moving in a crystal lattice will be subject to mutual interaction and to the influence of all the atoms at all the lattice sites. Around every atomic nucleus there is a Coulomb potential, which is inversely proportional to the distance from the atomic nucleus. This means that the electrons are situated in a potential that varies periodically with the lattice constant a (see fig. 2, which gives an example of such a potential in one dimension); the model of the electron behaviour has translational symmetry associated with the periodicity of the crystal lattice.

The state of an electron in a semiconductor can be described by the wave function $\psi(\mathbf{r})$, which is a solution of the Schrödinger equation (\mathbf{r} is the coordinate of position):

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}), \quad (1)$$

in which $\hbar = h/2\pi$, where h is Planck's constant, m is the electron mass, $\nabla^2 = (\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2)$, $\psi(\mathbf{r})$ is the wave function of the electron ($|\psi(\mathbf{r})|^2 d\mathbf{r}$ gives the probability of finding the electron between \mathbf{r} and $\mathbf{r} + d\mathbf{r}$), E is the energy of the electron corresponding to the state $\psi(\mathbf{r})$, and $V(\mathbf{r})$ is the crystal potential, i.e. the potential energy of the electron in the semiconductor material.

Solutions of eq. (1) are easily found in two extreme cases: when the potential $V(\mathbf{r})$ is very small and when the lattice constant a is very large. In the first case, when $V(\mathbf{r}) \approx 0$, we are dealing with free electrons, and the solutions $\psi(\mathbf{r})$ of the Schrödinger equation (1) consist of plane waves:

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (2)$$

which are characterized by the wave vector \mathbf{k} , or the momentum $\mathbf{p} = \hbar\mathbf{k}$. The corresponding wavelength is $\lambda = 2\pi/k$ (k is the length of \mathbf{k}) and the corresponding energy is given by

$$E(\mathbf{k}) = \frac{\mathbf{p}^2}{2m} = \frac{\hbar^2 \mathbf{k}^2}{2m}. \quad (3)$$

The energy values form a continuous spectrum as in fig. 3a. The probability of finding an electron between \mathbf{r} and $\mathbf{r} + d\mathbf{r}$, $|\psi(\mathbf{r})|^2 d\mathbf{r}$, is the same for any \mathbf{r} ; consequently, there is no preferential position for the electrons in a constant potential.

In the other case, when the distance between the atoms in the crystal lattice is very large, the electrons are bound to the individual atoms. The solution $\psi(\mathbf{r})$ of the Schrödinger equation for an individual atom is an atomic function characterized by the quantum numbers l and m , which are associated with the orbital angular momentum of the electron around the

nucleus. The energy spectrum is discrete (fig. 3*b*) and the energy levels are also characterized by the quantum numbers l and m . The probability of finding an electron is greatest in the vicinity of an atomic nucleus.



Fig. 2. Schematic representation of the crystal potential $V(x)$ encountered by electrons moving in a material with a perfect crystal lattice of lattice constant a .

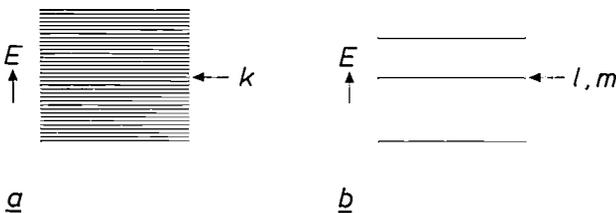


Fig. 3. Continuous spectrum of energy levels for a free electron (a) and a discrete spectrum of energy levels for a bound electron in an atom (b). The states of the free electron are characterized by the wave vector k ($\psi \propto e^{ik \cdot r}$). The states of the bound electron are characterized by the quantum number l of the orbital momentum and the magnetic quantum number m .

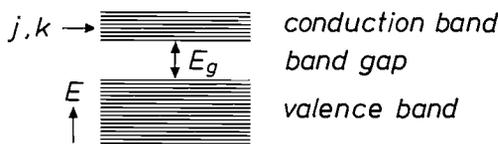


Fig. 4. Schematic representation of the energy levels of a semiconductor. States $\psi_{jk} = u_{jk} e^{ik \cdot r}$ are characterized by a wave vector k , which is related to the periodicity of the crystal lattice, and a quantum number j , which can be associated with the atomic origin of the function $u_{jk}(r)$. The difference in energy between the lowest conduction-band level and the highest valence-band level is the band gap E_g .

The energy spectrum of a semiconductor material lies between the two spectra discussed: it is neither continuous nor discrete, and consists essentially of a continuous part, the valence band, separated by a band gap E_g from a second continuous part of the spectrum, the conduction band (fig. 4). At low temperature all the valence-band states are occupied and all the conduction-band states are empty. The semiconductor then does not conduct electric current. At higher temperature there are some electrons in the conduction-band states and holes ('missing electrons') in the valence-band states. The semiconductor can now conduct an electric current. The electrons in the conduction band and the holes in the valence band

both act as carriers^[4]. When an electron returns from a conduction-band state to a valence-band state — a process known as electron-hole recombination — radiation is emitted at a wavelength determined by the energy difference between these two states.

Let us now take a closer look at the energy spectrum of a semiconductor. We have already seen that the states of free electrons (i.e. electrons that are not bound to a particular atomic nucleus) are characterized by a wave vector k and the states of the bound electrons in an atom are characterized by the quantum numbers l and m . Electrons in a crystal lattice are neither free nor bound to a particular atomic nucleus. Between the atomic nuclei the potential is relatively weak. Near an atomic nucleus the potential is relatively strong and the electron is then to some extent bound to that nucleus. So, in that picture, the solution of the Schrödinger equation in a semiconductor material is now seen to be composed of two parts: a contribution from free electrons and a contribution from a periodic function that describes the bond to the atomic nucleus. Fig. 5 shows this situation schematically for a one-dimensional potential. This solution may be written as

$$\psi_{jk}(r) = u_{jk}(r) e^{ik \cdot r} \quad (4)$$

The function u_{jk} consists essentially of the atomic wave functions at the positions of the atoms in the crystal lattice. The quantum number j is related to the atomic origin of $u_{jk}(r)$ and can therefore be expressed in terms of the quantum numbers l and m . The other quantum number, the wave vector k , is associated with the translational symmetry of the lattice. After translation through a distance R that is an integer number of times the lattice constant a in the one-dimensional example, the solution ψ must be such that the probability of finding the electron in a volume element dr remains unchanged; $|\psi(r + R)|^2 dr = |\psi(r)|^2 dr$. In other words, for each function $\psi(r)$ there is a vector k such that $\psi(r + R) = e^{ik \cdot R} \psi(r)$. This wave vector $k = (k_x, k_y, k_z)$ is thus characteristic of the solution

[3] R. Dingle, W. Wiegmann and C. H. Henry, Quantum states of confined carriers in very thin $Al_xGa_{1-x}As$ - $GaAs$ - $Al_xGa_{1-x}As$ heterostructures, Phys. Rev. Lett. **33**, 827-830, 1974; P. Dawson, G. Duggan, H. I. Ralph and K. Woodbridge, Free excitons in room-temperature photoluminescence of $GaAs-Al_xGa_{1-x}As$ multiple quantum wells, Phys. Rev. B **28**, 7381-7383, 1983; R. C. Miller and D. A. Kleinman, Excitons in $GaAs$ quantum wells, J. Lumin. **30**, 520-540, 1985; General information about layered semiconductors can be found in the special issue on 'Semiconductor quantum wells and superlattices: physics and applications', IEEE J. QE-22, No. 9, 1545-1922, 1986.

[4] To obtain sufficient conduction it is more usual to incorporate 'foreign' atoms in the lattice. Foreign atoms with an excess electron that they give up to the conduction band are called donors, those that have a 'missing' electron and take it from the valence band are called acceptors.

$\psi(\mathbf{r})$. It can be seen in fig. 5 that $\psi_{j\mathbf{k}}$ does not vary periodically with the lattice constant a , but that the probability $|\psi_{j\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r}$ (which is equal to $|u_{j\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r}$ since $|e^{i\mathbf{k}\cdot\mathbf{r}}| = 1$) of finding an electron between \mathbf{r} and $\mathbf{r} + d\mathbf{r}$ does vary in this way.

If the wavelength $\lambda = 2\pi/k$ is much larger than a , then $e^{i\mathbf{k}\cdot\mathbf{r}}$ varies much more slowly as a function of \mathbf{r} than $u_{j\mathbf{k}}(\mathbf{r})$. The free-electron function $e^{i\mathbf{k}\cdot\mathbf{r}}$ can then

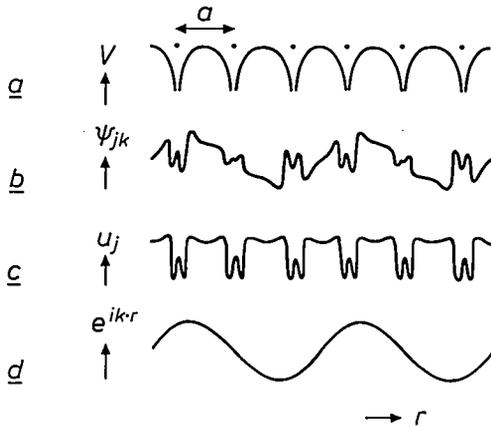


Fig. 5. a) Schematic representation of the crystal potential $V(\mathbf{r})$ encountered by electrons moving in a material with a perfect crystal lattice of lattice constant a . b) Example of a wave function $\psi_{j\mathbf{k}}$. This can be described by the product of a function $u_{j\mathbf{k}}(\mathbf{r})$, periodic in the lattice constant a and approximating to a combination of the atomic wave functions at the different atomic positions in the lattice (c), and a plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$, the solution for a free electron (d).

be considered as an envelope of the solution $\psi_{j\mathbf{k}}$ in the semiconductor. The function $u_{j\mathbf{k}}(\mathbf{r})$ contains the information about the bonds between the electrons and the atoms of the semiconductor material.

In a semiconductor material we are interested in emission (or absorption) of radiation when electrons go from the conduction band to the valence band or vice versa. For $\lambda \gg a$ the functions $u_{j\mathbf{k}}(\mathbf{r})$ corresponding to the states with the highest energies of the valence band, the top of the valence band, are found to consist of the atomic p_x , p_y and p_z states of the individual atoms in the semiconductor material, for which $l = 1$ and $m = 0, -1, +1$ respectively. The functions $u_{j\mathbf{k}}(\mathbf{r})$ corresponding to the states with the lowest energies of the conduction band, the bottom of the conduction band, are found, to an approximation, to consist of atomic s states of the individual atoms, for which $l = 0$ and $m = 0$. This therefore gives one conduction-band state and three valence-band states for each wave vector \mathbf{k} . Fig. 6 shows the part of the energy spectrum, E as a function of $\mathbf{k} = (k_x, 0, 0)$, in which these states occur for GaAs.

The bottom of the conduction band, which we shall refer to as $j = e$ (for 'electron') can be described approximately by a 'free-electron band':

$$E_e(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m_e^*} + E_g \tag{5}$$

The mass m_e^* in this equation, the 'effective electron mass' [6], is not the mass of the free electron, but is much smaller ($m_e^* = 0.07 m_e$). An electron in a semiconductor material does not, of course, behave as a really free electron because it is always affected by the individual atoms. The influence of the lattice is taken into account by describing the movement of the electron as that of a free electron of smaller mass.

The top of the valence band consists of three separate bands associated with the p states in the individual atoms. Two of these are designated by hh (for 'heavy hole') and lh (for 'light hole'). The third band is split off because of the coupling between the orbital momentum and the spin of the electrons, and is designated so (for 'split off'). In the same way as for the electrons in the conduction band, we can describe the behaviour of the holes by means of the equations for the free holes with effective hole masses:

$$E_{hh}(\mathbf{k}) = - \frac{\hbar^2 \mathbf{k}^2}{2m_{hh}^*},$$

$$E_{lh}(\mathbf{k}) = - \frac{\hbar^2 \mathbf{k}^2}{2m_{lh}^*},$$

$$E_{so}(\mathbf{k}) = - \Delta - \frac{\hbar^2 \mathbf{k}^2}{2m_{so}^*} \tag{6}$$

This applies only when \mathbf{k} is sufficiently close to zero. At slightly higher values of the wave vector \mathbf{k} , the description is less exact, owing to the interaction of the holes in these energy states (in fig. 6 it can be seen that the lh and the so bands 'repel' one another).

We saw above that holes associated with the various p orbitals in the atoms were located in different bands and had different effective masses. This can be explained as follows. Fig. 7 shows the electron distribution, which is proportional to $|\psi(\mathbf{r})|^2 d\mathbf{r}$, for the p_x , p_y and p_z orbitals around two atoms in the lattice. If the electron distributions around the atoms overlap, the electrons can move easily through the lattice from one atom to another. Electrons (and hence holes) in the p_x orbitals can therefore move easily through the lattice in the x -direction. These holes are the light holes. Electrons (and holes again) in the p_y and p_z orbitals cannot move so easily from atom to atom in the x -direction through the lattice. Holes in

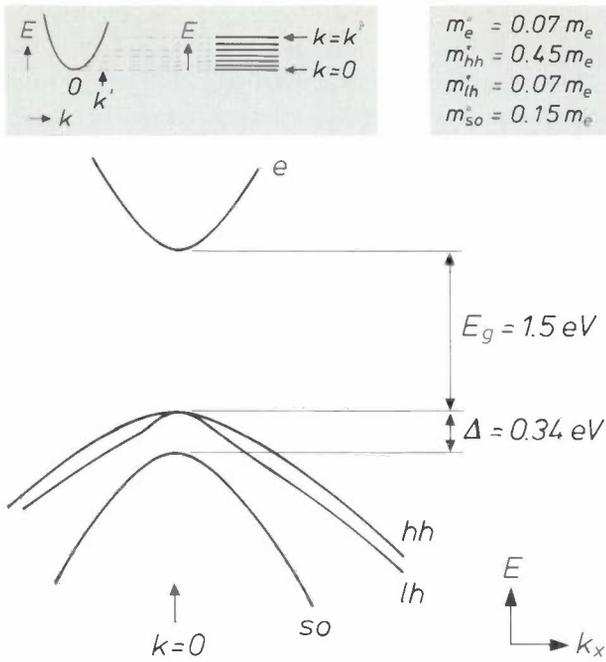


Fig. 6. Band structure $E(k)$ of GaAs for $|k| \ll 2\pi/a$ and k parallel to the [100] crystal axis of the zinc-blende GaAs lattice. Only the principal bands around the band gap are indicated: the conduction band, denoted by e, and the three valence bands. The energy-difference between the top of both the heavy-hole band (hh) and the light-hole band (lh) and the spin-orbit split-off hole band (so) is Δ . The effective masses m_j^* of the bands are expressed in terms of the electron mass m_e . The relationship between this representation and fig. 3 and fig. 4 is illustrated at the top of the figure.

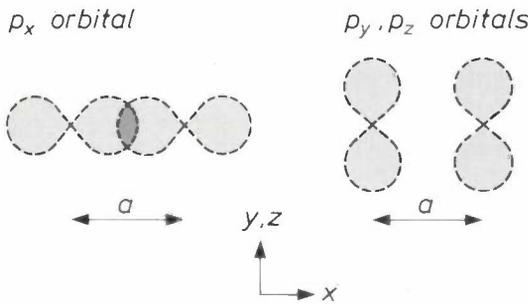


Fig. 7. Schematic representation of the electron distribution associated with the three p orbitals. For electrons moving in the x -direction ($k = (k_x, 0, 0)$) the p_x -functions correspond to the light holes and the p_y -functions and p_z -functions to the heavy holes.

these orbitals will therefore behave less like free holes, and consequently have larger effective masses. These are the heavy holes.

Equations (5) and (6) for the energy spectrum closely resemble eq. (3) for a free electron. They apply only near $k = 0$. A good approximation for the wave function $\psi_{j\mathbf{k}}$ near $k = 0$ is therefore:

$$\psi_{j\mathbf{k}}(\mathbf{r}) = u_{j\mathbf{k}=0} F_{j\mathbf{k}}(\mathbf{r}), \quad (7)$$

where $F_{j\mathbf{k}}(\mathbf{r})$, the envelope function, takes on the same

role as $e^{i\mathbf{k}\cdot\mathbf{r}}$ in eq. (4): the envelope function is a slowly varying function on the scale of the lattice constant. Further manipulation of the Schrödinger equation (1) brings us to equation (8) for the envelope function:

$$\left[-\frac{\hbar^2}{2m_j^*} \nabla^2 + E_j \right] F_{j\mathbf{k}}(\mathbf{r}) = E F_{j\mathbf{k}}(\mathbf{r}), \quad (8)$$

where E_j stands successively for $E_e = E_g, E_{hh} = E_{lh} = 0$ and $E_{so} = -\Delta$. Solutions of this equation are of the free-electron kind (2), i.e. $F_{j\mathbf{k}} \propto e^{i\mathbf{k}\cdot\mathbf{r}}$. The corresponding energy levels are indeed given by equations (5) and (6). The envelope function $F_{j\mathbf{k}}$ varies slowly on the scale of the lattice constant and it is therefore not surprising that the crystal potential $V(\mathbf{r})$ no longer appears explicitly in eq. (8).

We shall give the main steps in the derivation of equations (7) and (8) without going into the details [6]. We expand $u_{j\mathbf{k}}(\mathbf{r})$, the periodic part of the solution of Schrödinger's equation $\psi_{j\mathbf{k}}(\mathbf{r})$, into periodic functions $u_{i\mathbf{k}=0}(\mathbf{r})$ (a complete orthonormal set of solutions of the Schrödinger equation at $k = 0$; i goes over all the bands). The result can be expressed in the form:

$$\psi_{j\mathbf{k}} = \sum_i F_{i\mathbf{k}}(\mathbf{r}) u_{i\mathbf{k}=0}(\mathbf{r}), \quad (a)$$

where $F_{i\mathbf{k}}$ are the envelope functions. After inserting this expansion in the Schrödinger equation (1) we recognize the Schrödinger equation for $u_{i\mathbf{k}=0}(\mathbf{r})$, and insert the corresponding energy values E_i . We multiply the remaining equation by $u_{i\mathbf{k}=0}^*(\mathbf{r})$, the complex conjugate of $u_{i\mathbf{k}=0}(\mathbf{r})$, from the left. After integration over the unit cell we are left with:

$$\left(\frac{-\hbar^2}{2m} \nabla^2 + E_j \right) F_{j\mathbf{k}} + \sum_{i \neq j} \frac{\hbar}{im} \mathbf{p}_{ji} \cdot \nabla F_{i\mathbf{k}} = E_j F_{j\mathbf{k}}, \quad (b)$$

where $\mathbf{p}_{ij} = \int d\mathbf{r} u_{i\mathbf{k}=0}^*(\mathbf{r}) (\hbar/i) \nabla u_{j\mathbf{k}=0}(\mathbf{r})$. A result of integrating over the unit cell is that the crystal potential $V(\mathbf{r})$ no longer appears as a separate term in the equation. Instead we now have a set of simultaneous equations for the envelope functions $F_{j\mathbf{k}}(\mathbf{r})$. The link between the various equations is the momentum operator $\mathbf{p} = (\hbar/i) \nabla$.

We shall now only consider values of k close to zero. The envelope function now varies slowly on the scale of the lattice constant. The main contribution to the solution $\psi_{j\mathbf{k}}$ for each band j for small values of k will come from the envelope function $F_{j\mathbf{k}}(\mathbf{r})$ (the main contribution to the corresponding energy $E_{j\mathbf{k}}$ comes from E_j). Since we are interested in the relative positions of the energy levels (i.e. in

[5] The effective electron mass is defined as $m^* = F/a$, where F is the external force acting on the electron and a is the acceleration of the electron as a result of the sum of the external force and the forces due to the interaction with the lattice. Bearing in mind that $a = dv/dt = (dv/dk)(dk/dt)$, $F = (\hbar dk/dt)$ (where v is the velocity of the electron) and $v = (dE/dk)/\hbar$, then we have $m^* = \hbar^2 (d^2E/dk^2)^{-1}$.

[6] More details of this derivation are given in: E. O. Kane, Energy Band Theory, in Handbook on Semiconductors, North-Holland, New York, 1982, Vol. 1, pp. 193-217 and in R. Eppenga, M. F. H. Schuurmans and S. Colak, A new k.p theory for GaAs/Ga_{1-x}Al_xAs-type quantum wells, Phys. Rev. B 36, 1554-1564, 1987.

the differences $E_{j\mathbf{k}} - E_j$) we cannot just neglect the effects of the other bands. The effects of the band coupling can be taken into account by a perturbation method. The final equation then becomes:

$$\left(\frac{-\hbar^2}{2m_j^*} \nabla^2 + E_j \right) F_{j\mathbf{k}} = E_{j\mathbf{k}} F_{j\mathbf{k}}, \quad (c)$$

where

$$\frac{1}{m_j^*} = \frac{1}{m} + \frac{1}{m^2} \sum_{i \neq j} \frac{p_{ji} \cdot p_{ij}}{E_j - E_i}.$$

The second term in this equation thus contains the interaction with all the other bands. The influence of the crystal potential, which we described first by the set of simultaneous equations for the envelope functions, is now contained in a single parameter, the effective mass m_j^* .

The free quasiparticle described by equation (8) for band j , known as the Bloch electron or Bloch hole for that band, is an 'electron' or 'hole' in which the interaction with the lattice and all the other electrons is taken into account: the mass of the Bloch electron is not the free-electron mass but the effective electron mass. In an external potential that varies slowly on the scale of the lattice constant we can describe the movement of the Bloch particles by adding that potential to equation (8). A Bloch electron in an external potential is accelerated as if it were a free electron with a different mass, i.e. the effective mass.

Now that we have gained some understanding of the behaviour of charge carriers in a semiconductor material, we can return to the actual problem, the GaAs/AlGaAs quantum well. Since the lattice constants in both materials are virtually identical, these materials can be applied to one another in a perfect lattice match. The movement of the charge carriers, however, will be affected in a different way by the different atoms.

The locations of the energy levels in the valence and conduction bands are shown in *fig. 8a* for the bulk materials that constitute the quantum-well structure. Since the conduction band in GaAs is lower and the valence band higher than the corresponding bands in AlGaAs, electrons and holes in GaAs are more strongly bound to the lattice, and in this material there is consequently a larger concentration of charge carriers and hence a greater probability of recombination. In a GaAs layer between two AlGaAs layers, as in *fig. 8b*, the movement of electrons is restricted by a rectangular potential well that is formed as a result of the energy difference between the conduction bands of GaAs and AlGaAs. In the same way the movement of the holes is restricted by a rectangular potential well

as a result of the energy difference between the valence bands of GaAs and AlGaAs.

This formation of potential wells for the electrons and holes as a result of the energy differences between the levels in AlGaAs and those in the GaAs layer is not of trivial significance. It is justified by the envelope-function formalism, since the averaging removes the microscopic details. If the thickness of the GaAs layer is many lattice constants, then equation (8) is applicable to both the thin GaAs layer and the thick AlGaAs layers. The effective masses m_j^* and the locations of the energy levels E_j in the two materials are different, of course: we give these quantities different superscripts to distinguish them from one another in the different materials. We put m_j^a and E_j^a for the well material and m_j^o and E_j^o for the barrier material confining the well. The resulting equations can be comprised in a single equation:

$$\left[- \frac{\hbar^2}{2m_j^*(x)} \nabla^2 + V_j(x) \right] F_j(\mathbf{r}) = E F_j(\mathbf{r}), \quad (9)$$

where $j = e, hh, lh$, so and with $m_j^*(x) = m_j^a$ and $V_j(x) = E_j^a$ for $|x| < w/2$, and $m_j^*(x) = m_j^o$ and $V_j(x) = E_j^o$ for $|x| > w/2$ (w is the thickness of the GaAs layer). It is found that $E_e^a < E_e^o$ (see *fig. 8*), in other words, electrons prefer to be in the well material GaAs rather than in the barrier material AlGaAs. Except for the position-dependent effective mass, equation (9) thus describes the confinement of a Bloch electron in a rectangular potential well. The depth of the well is $\Delta V_e = E_e^o - E_e^a$, the energy difference between the conduction bands of AlGaAs and GaAs, and the width w of the well is equal to the thickness of the GaAs layer. For the holes it is found that $E_h^a > E_h^o$, in other words it requires less energy to make a hole in the well material than in the barrier material. Bloch holes are therefore also confined in a rectangular potential well of depth $\Delta V_h = E_h^o - E_h^a$, and for the light and heavy holes this corresponds to the energy difference between the valence bands of GaAs and AlGaAs.

A description of the permissible states for the carriers therefore requires little more than a solution of the well-known quantum-mechanical problem of a particle in a box. We say 'little more' here because the difference in effective masses in the barrier and well materials enters into the argument. The problem is also a little more complicated than appears from eq. (9). This equation describes non-coupled electron and hole bands. In reality, however, there is some coupling, and the coupling between the hole bands is in fact quite considerable. This coupling is visible in the band structure of the bulk material GaAs. In *fig. 6* it can be seen that the light-hole band and the spin-

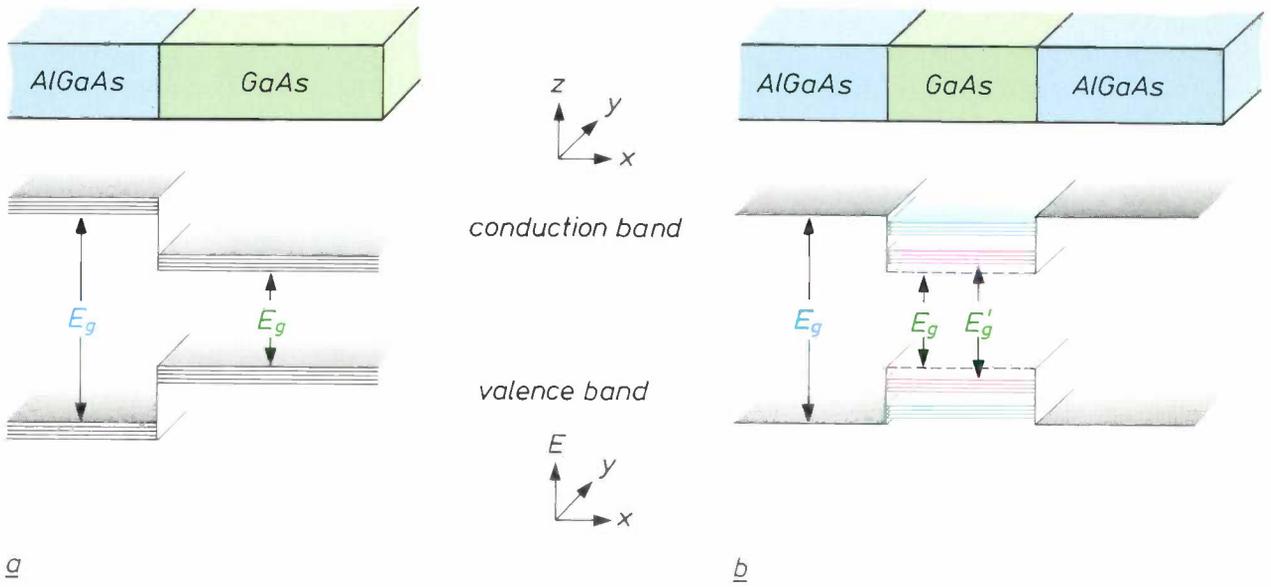


Fig. 8. *a)* Schematic representation of the locations of the energy levels at the interface between AlGaAs and GaAs. The difference in energy between the valence band and the conduction band is greater for AlGaAs than for GaAs. At low temperature only the valence band is filled, at higher temperature, or when a voltage is applied, there are electrons in the conduction band as well. *b)* Locations of the energy levels in a quantum-well structure, a thin layer (about 10 nm) of GaAs between two thicker layers of AlGaAs. The movement of carriers in the *x*-direction is confined, which leads to quantization of their energy. The difference in energy between the conduction band and the valence band is greater in the thin layer of GaAs than the corresponding difference in the bulk material ($E'_g > E_g$).

orbital split-off band ‘repel’ one another (at the point where the hole bands differ from parabolic curves the light-hole band and the spin-orbital split-off band separate) — another effect not included in the description of equation (6).

In the next section we shall discuss the problem of a particle in a quantum well, starting from the description of a particle in a box. The effects of the refinements mentioned above on the ‘envelope-function formalism’ will not be considered further here, but we shall show some of the results of our calculations with the modified formalism.

A particle in a quantum well

In this section we shall adopt a step-by-step approach to the problem of a particle in a quantum well, described by equation (9). First, we shall solve the problem for a particle in a one-dimensional potential well (the particle-in-a-box problem). We shall then extend the problem to three dimensions and finally we shall consider the particle in the quantum well.

The movement of a particle that can move freely between $x = -w/2$ and $x = +w/2$ (a one-dimensional potential well, see fig. 9) can be described by the sol-

ution $\psi(x)$ of the one-dimensional Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E\psi(x). \quad (10)$$

For an infinitely deep well (fig. 9*a*) the only allowed solution outside the well, ($x < -w/2$ and $x > +w/2$), is $\psi(x) = 0$. Inside the well there are solutions of the form $\psi(x) \propto \cos kx$ and $\psi(x) \propto \sin kx$ where $k = \sqrt{2mE}/\hbar$. For a well of finite depth (fig. 9*b*) the same solutions $\psi(x) \propto \cos kx$ and $\psi(x) \propto \sin kx$ remain valid inside the well. Outside the well, however, the solution now takes the form $\psi(x) \propto e^{\pm\kappa x}$. Since the value of the wave function now differs from zero outside the potential well, there must be some probability of finding the particle outside the potential well. The wave functions $\cos kx$ inside the well $e^{\pm\kappa x}$ and outside it constitute a solution of the Schrödinger equation for the same value of the energy E if k and κ are given by $k = \sqrt{2mE}/\hbar$ and $\kappa = \sqrt{2m(\Delta V - E)}/\hbar$.

The wave function cannot change its value abruptly, even at the edges of the potential well: the wave function $\psi(x)$ and its derivative must be continuous everywhere. From this condition and the boundary

conditions that $\psi(x) \rightarrow 0$ for $x \rightarrow \pm \infty$ it follows that solutions can only be found for certain values of k and κ . Since the energy of the particle is directly related to k and κ , the possible values for the energy are therefore fixed.

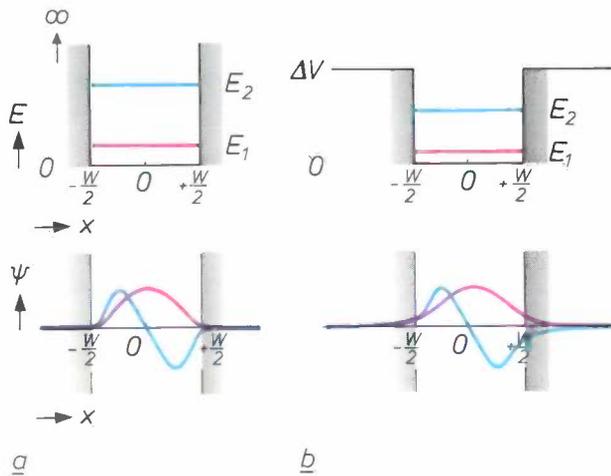


Fig. 9. The first two solutions ($n = 1, 2$; energy E and wave function $\psi(x)$) of the Schrödinger equation for the problem of a particle in an infinitely deep one-dimensional potential well (a) and a one-dimensional potential well of finite depth (b). The width of the well is w .

For the infinitely deep well it follows from the boundary conditions that $k = n\pi/w$ and the allowed energies are $E_n = (\hbar^2/2m)(n\pi/w)^2$. (For $n = 1, 3, 5, \dots$ this corresponds to a symmetrical solution ($\psi(x) = \psi(-x)$), and for $n = 2, 4, 6, \dots$ it corresponds to an antisymmetrical solution ($\psi(x) = -\psi(-x)$.) The locations of the energy levels and the wave functions for the first two solutions ($n = 1, 2$) are indicated in fig. 9.

For a well of finite depth the allowed values of the energy can be calculated numerically from the equations that follow from the boundary conditions: $\tan(kw/2) = \kappa/k$ for the symmetrical solution and $\cot(kw/2) = -\kappa/k$ for the antisymmetrical solution. We find that the allowed energy levels of a potential well of finite depth are lower than the corresponding levels of an infinitely deep potential well.

Solutions of the Schrödinger equation for a particle that can move freely in two dimensions but whose movement in the third dimension is confined by a potential well of finite depth are found in the product of the wave function for the one-dimensional particle-in-a-box problem and the wave function for a particle that can move freely in the two other dimensions: $\psi(\mathbf{r}) = e^{i\mathbf{k}_{\parallel} \cdot \mathbf{r}_{\parallel}} \cos k_n x$, where $\mathbf{r}_{\parallel} = (y, z)$, $\mathbf{k}_{\parallel} = (k_y, k_z)$ and where k_n relates to the allowed energies E_n . The total energy of the particle is characterized by the quantum number n and \mathbf{k}_{\parallel} and is given by:

$$E_n(\mathbf{k}_{\parallel}) = \frac{\hbar^2 \mathbf{k}_{\parallel}^2}{2m} + \frac{\hbar^2}{2m} \left(\frac{n\pi}{w} \right)^2 \quad (n = 1, 2, 3, \dots) \quad (11)$$

The energy thus contains a contribution that increases continuously as a function of \mathbf{k}_{\parallel} , and a discrete contribution. Fig. 10 shows the total energy as a function of \mathbf{k}_{\parallel} .

The electrons and holes in the quantum well behave like particles in a box. Their movement in the x -direction is confined by the relative positions of the energy levels in GaAs and AlGaAs. Their behaviour is described by eq. (9). The effective masses in this equation are functions of the position x : the effective mass is different in the well material and the barrier material. The solution of this equation with the appropriate boundary conditions (continuous wave function, so that $F_j(x)$ and $(\partial F_j / \partial x) / m_j^*$ must be continuous for the envelope) gives:

$$E_{jn}(\mathbf{k}_{\parallel}) = \frac{\hbar^2 \mathbf{k}_{\parallel}^2}{2m_j^*} + E_{jn}, \quad (12)$$

where $j = e, hh, lh$, and so, $n = 1, 2, \dots$. The depth of the potential well is different for the different carriers: $\Delta V_j = E_j^o - E_j^q$. These energy levels are fixed by

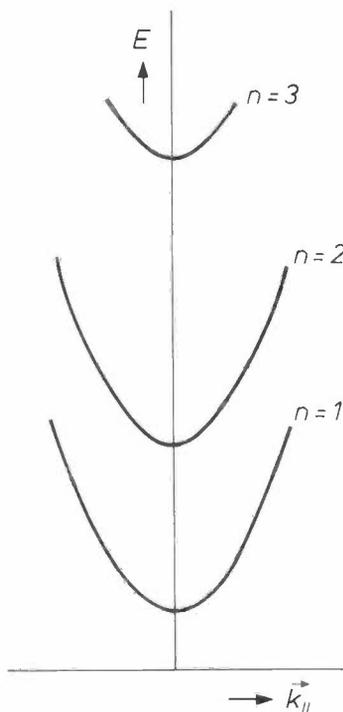


Fig. 10. Energy E as a function of \mathbf{k}_{\parallel} for a particle in a three-dimensional potential well (see eq. 11). The movement of the particle in the x -direction is limited, but it can move freely in the y - and z -directions.

the equations that follow from the boundary conditions:

$$k = \sqrt{\frac{2m_j^a E}{\hbar^2}}, \quad \kappa = \sqrt{\frac{2m_j^o(\Delta V_j - E)}{\hbar^2}}, \quad \tan \frac{k w}{2} = \frac{m_j^a \kappa}{m_j^o k} \quad (13a)$$

for the symmetrical solutions and

$$k = \sqrt{\frac{2m_j^a E}{\hbar^2}}, \quad \kappa = \sqrt{\frac{2m_j^o(\Delta V_j - E)}{\hbar^2}}, \quad \cot \frac{k w}{2} = -\frac{m_j^a \kappa}{m_j^o k} \quad (13b)$$

for the antisymmetrical solutions.

Solutions of these sets of equations are given by the dashed curves in *fig. 11* for the electrons and in *fig. 12* for the holes. These figures show the locations of the energy levels and the total density of states, i.e. the number of states per unit energy and volume.

The results for the electron bands (*fig. 11*) show considerable resemblance to *fig. 10*: the electron bands are parabolic functions of k_{\parallel} . The density of states as a function of the energy is a stepped function. As soon as the energy reaches the bottom of the next band, more energy levels are added and the density of states increases by a step.

The picture is rather more complicated for the holes (*fig. 12*). The holes in the so band are not shown in this figure (the energy of these holes lies outside the potential well for the so holes). For the light and the

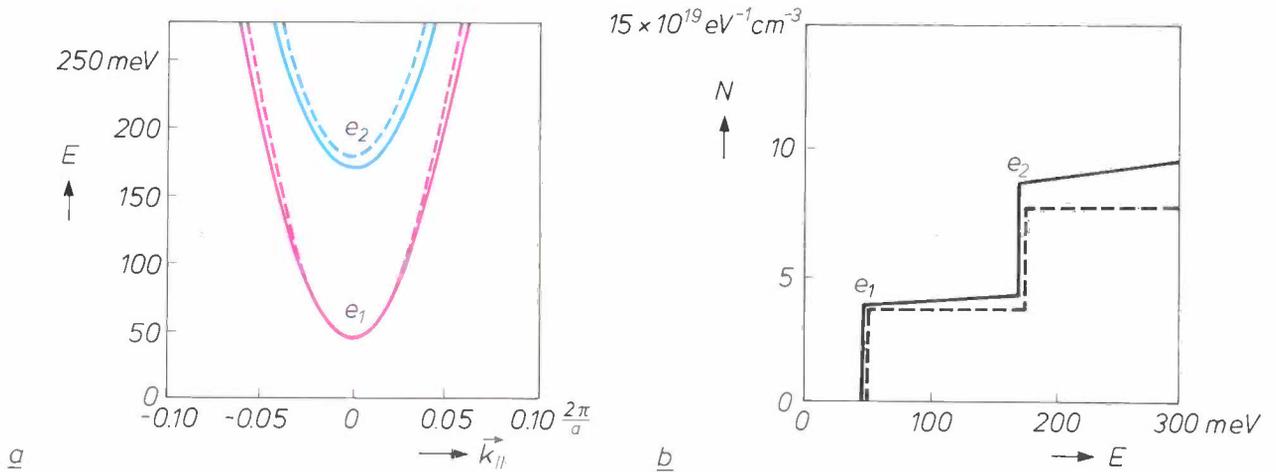


Fig. 11. The calculated band structure, *a*) the energy E as function of k_{\parallel} and *b*) the corresponding density of states for the electron bands with $n = 1$ (red) and $n = 2$ (blue) in a quantum well consisting of a 7.7-nm layer of GaAs between two thicker layers of $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$. The dashed curves give the results with no band mixing, the continuous curves show the results where band mixing is taken into account.

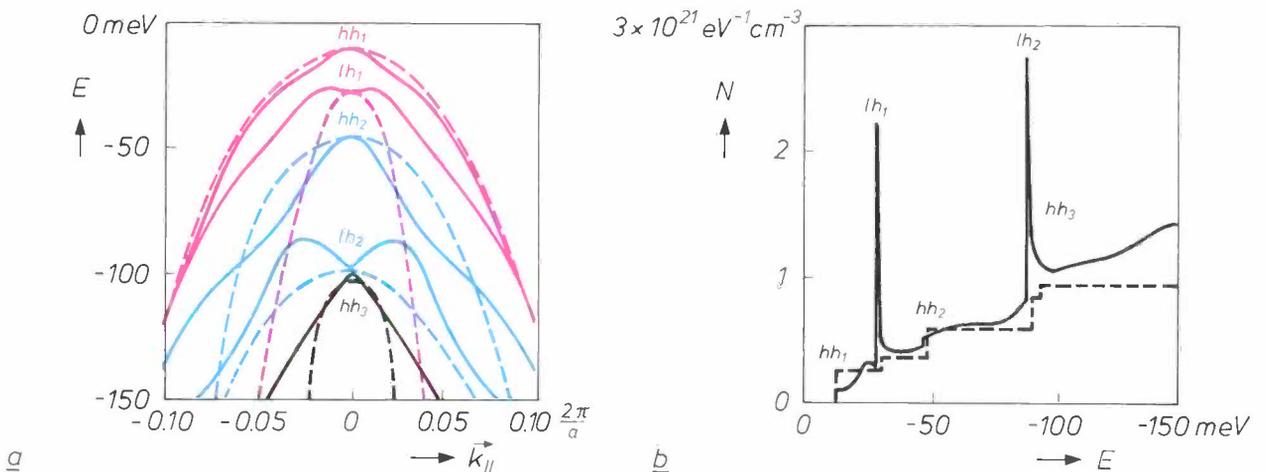


Fig. 12. Calculated band structure for the hole bands. The notation is as in *fig. 11*.

heavy holes fig. 12 gives the parabolic energy bands for $n = 1, 2, 3$. In the quantum well the tops of the heavy-hole and light-hole bands do not coincide, as they do in the bulk material. The density of states is a stepped function of the energy: every time the upper level of a new band is crossed, the density of states increases by a step. (The size of the steps is different for the heavy-hole and light-hole bands.)

It is immediately obvious that the hole bands cross each other — an unacceptable situation physically. At the $k_{//}$ -values at which crossings of the hole bands occur the interaction between the bands is exceptionally strong. When this interaction — ‘band mixing’ — is included in the calculations, simultaneous equations are obtained for the envelope function. Numerical solution of these equations gives the results given by the continuous lines in fig. 11 and fig. 12. Including band mixing in the calculations has particularly important consequences for the hole bands. The bands no longer cross and the density of states is very different. The peaks in the density of states function occur when $dE/dk_{//}$ becomes zero for $k_{//} \neq 0$ (see fig. 12).

Applications of the theory

Absorption spectrum of a quantum well

In this section we shall discuss a result obtained by calculations based on the theory of the quantum well as outlined above. It relates to the absorption spectrum of a quantum well consisting of a 7.7-nm layer of GaAs between two thick layers of $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$. The quantum well has been grown in the [001] crystal direction. We shall see that the calculated absorption is in good agreement with the measured absorption spectrum^[7] of this quantum well.

Radiation of energy $\hbar\omega$ is absorbed by the quantum well structure if $\hbar\omega$ corresponds to the energy difference between two levels with an allowed transition between them. (In a quantum well, just as in an atom, there are allowed and forbidden transitions between energy levels.) Fig. 13 shows a calculated absorption spectrum for a quantum well. The dashed line represents the simplest case, with no account taken of band mixing. The absorption as a function of the energy $\hbar\omega$ of the radiation is a stepped function: whenever the energy crosses the bottom of the next band, more transitions are possible. The inclusion of band mixing in the calculations results in the spectrum given by a continuous line. Just as for the density of hole states, there are now peaks on the stepped function.

Another effect that comes into the calculations besides band mixing is the Coulomb interaction. A hole in the valence band and an electron in the conduction band attract one another, forming a kind of hydrogen

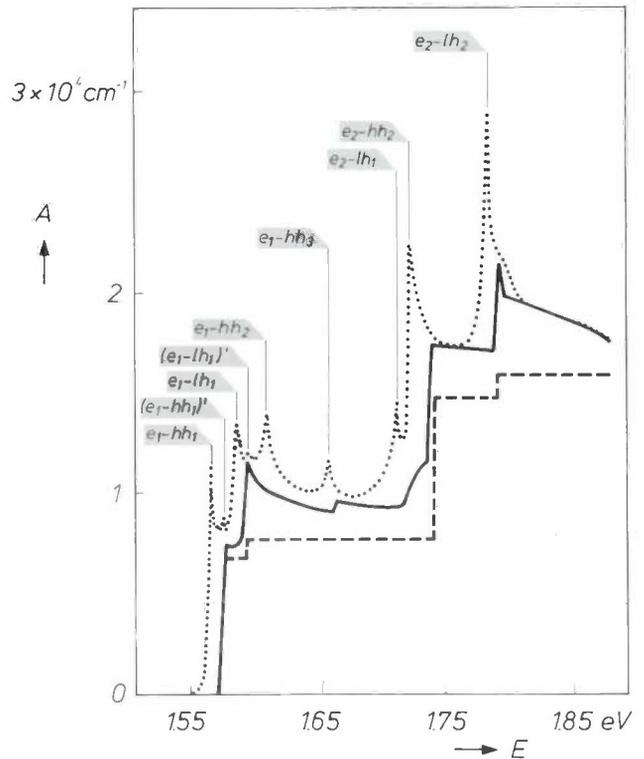


Fig. 13. Calculated absorption spectrum, absorption A as a function of the energy E of the radiation, for a quantum well consisting of a 7.7-nm thick layer of GaAs between two thicker layers of $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$. The dashed curve gives the results of the calculations without band mixing, the solid curve the results of calculations with band mixing, and the dotted curve indicates the results of calculations with band mixing, but now with the energy state of the resulting exciton also taken into account. The positions of the extra peaks in the spectrum due to the inclusion of the exciton energy state have been calculated. The shapes of the peaks are only approximated. The labels by the peaks specify the state of the exciton after absorption of the radiation (the notation for the transition is given in the text).

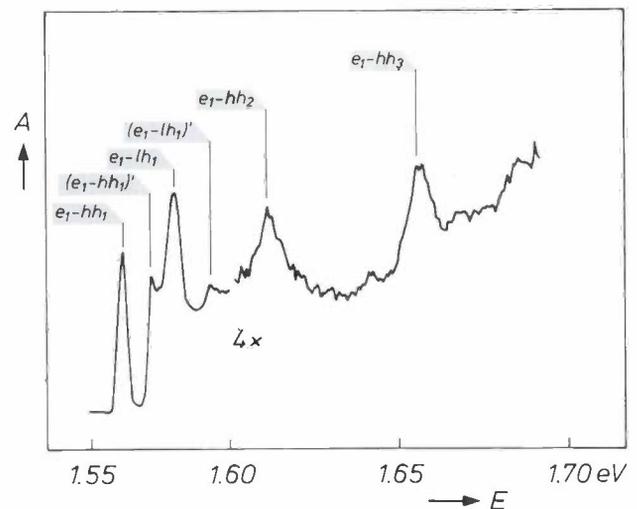


Fig. 14. Measured absorption spectrum, absorption A as a function of the energy E of the radiation of a quantum well (7.7-nm layer of GaAs between $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$). The arrows indicate the calculated energies of the peaks in the spectrum as a result of the transition shown (the notation for the transitions is given in the text). Above $E = 1.6$ eV the absorption scale is increased by a factor of 4.

atom. As in hydrogen atoms, the energy of these bound electrons and holes — called excitons — can assume discrete values that depend on the distance between hole and electron. This can be included in the calculation by slightly shifting the energy levels in the quantum well towards lower energy values. When this is done in the calculations of the absorption spectrum, a curve rather like the dotted curve in fig. 13 is obtained. A number of extra peaks appear in the absorption spectrum, and these can be attributed to tran-

sitions between the different bands. The same excellent agreement is found for quantum wells that have been grown in a different direction.

Gain spectrum of a quantum-well laser

Another practical application of this theory is in modelling the optical characteristics of GaAs/AlGaAs quantum-well lasers. Quantum-well lasers are layered semiconductor structures that contain one or more quantum wells. In the design of laser structures the

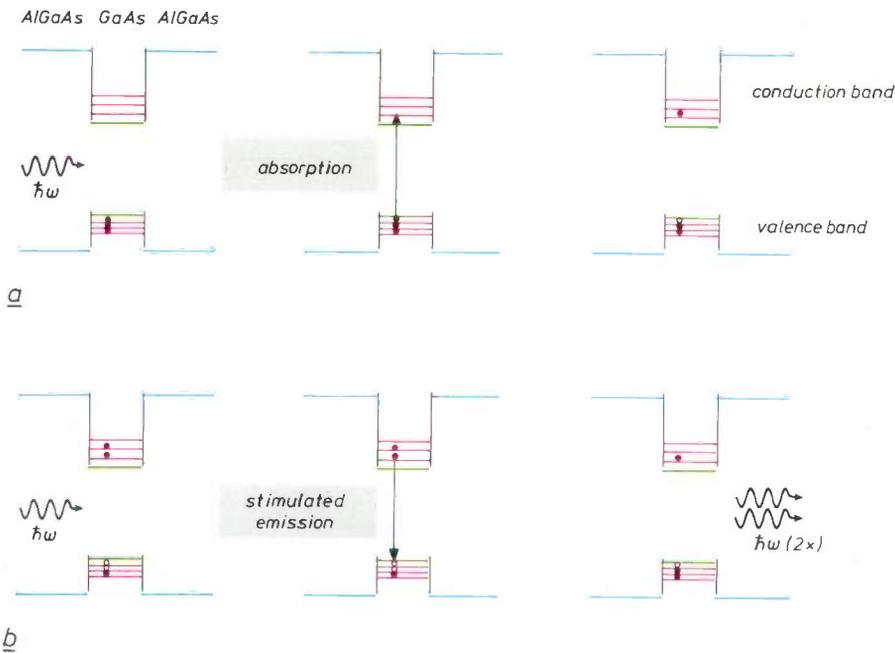


Fig. 15. Schematic representation of the transitions on absorption (a) of a photon and stimulated emission (b) arising from a photon of energy $\hbar\omega$. For clarity only a few levels are indicated in the conduction band and the valence band in the quantum well. Levels occupied by electrons are indicated by a dot (•). A 'missing' electron in a valence-band level is indicated by a small circle (○). When a photon of energy $\hbar\omega$ is absorbed an electron goes from a valence-band state to a conduction-band state; in stimulated emission a photon of energy $\hbar\omega$ causes an electron to return from the conduction-band state to a valence-band state.

sitions between the different bands. (For instance, the transition between the first electron band e_1 and the first heavy-hole band hh_1 , is denoted by $e_1 - hh_1$, whereas the same transition but resulting in a different state of the exciton is designated by $(e_1 - hh_1)'$.)

The absorption spectrum calculated in this way (with band mixing and excitons included) can be compared with experimental spectra. This can be seen in fig. 14. The arrows in the figure show where the calculations indicate there ought to be a peak in the spectrum. In fact absorption peaks are found at the expected places in the experimental spectrum. The agreement between the measured and the calculated spectra

composition of the layered structures and the dimensions of the different layers are chosen in such a way that the principal characteristics of the laser, the wavelength and the threshold current, satisfy specific conditions. It is therefore important that the relation between the parameters of the layered structure and the final characteristics of the laser is known. This relation can of course be found by making measurements on finished lasers, but this is time-consuming and ex-

[7] L. W. Molenkamp, R. Eppenga, G. W. 't Hooft, P. Dawson, C. T. Foxon and K. J. Moore, Determination of valence-band effective-mass anisotropy in GaAs quantum wells by optical spectroscopy, submitted for publication to Phys. Rev. B (1988).

pensive. Detailed models for the process in which laser light is generated by recombination of holes and electrons in the quantum well can be of great value in the rapid design of an optimum laser structure.

In essence a quantum-well laser consists of an n-doped AlGaAs layer, an undoped (intrinsic) region consisting of AlGaAs with a lower Al content and containing a number of thin GaAs layers (quantum wells), and a p-doped AlGaAs layer. This combination of a p-type layer, an n-type layer and an intrinsic region between them is called a pin junction. The cleavage planes at the ends of the layered structure act as reflectors to provide the required optical confinement. An appropriate voltage across the pin junction causes an electron current to flow from the n-doped region and a hole current to flow from the p-doped region to the intrinsic region. This produces inversion in the intrinsic region: some levels at the bottom of the conduction band are occupied by electrons and in some levels at the top of the valence band electrons are 'missing', which means that the latter levels are occupied by holes. The electrons and holes in the intrinsic region 'fall' into the quantum wells and recombine there, emitting photons in the process. If the current is large enough, laser action is produced because of the optical confinement.

Prediction of the wavelength of the laser light and the threshold current of the laser requires a very accurate knowledge of the gain spectrum and the spontaneous emission spectrum of the GaAs/AlGaAs quantum well. In calculating the emission spectrum, just as in calculating the absorption spectrum, we need to know whether a transition between two levels is allowed or not. Besides this, there is a further condition that determines the nature of the transition; the occupancy of the levels determines whether there is absorption or emission. Absorption of a photon can only arise between an occupied valence-band level and an unoccupied conduction-band level; emission of a photon can only arise between an occupied conduction-band level and an unoccupied valence-band level (see *fig. 15*). The occupation probability of the levels as given by the Fermi-Dirac distribution can be calculated if the temperature of the crystal lattice and the number of injected electrons and holes are known. *Fig. 16* shows a gain spectrum calculated taking level occupancy into account (continuous line). The gain is zero for photon energies less than the band gap. At higher photon energies there is gain (by stimulated emission) arising from transitions between occupied levels at the bottom of the conduction band and unoccupied levels at the top of the valence band. At even higher photon energies the gain finally becomes negative, i.e. there is absorption because of transitions be-

tween occupied valence-band levels and unoccupied conduction-band levels.

For a comparison between calculated and measured spectra other effects have to be taken into account as

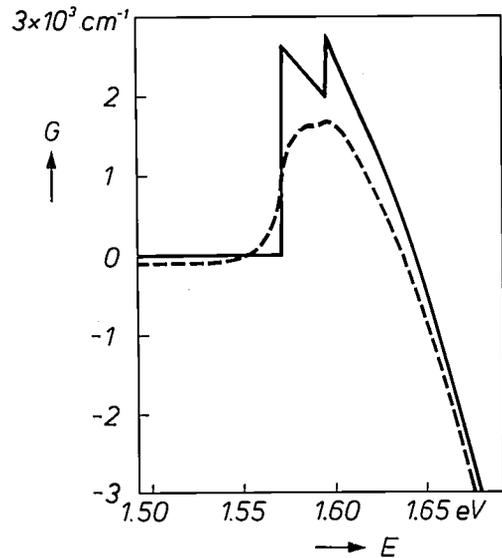


Fig. 16. Calculated gain spectrum. Gain G as a function of the energy E of the radiation for a quantum-well laser in which there is a layer of GaAs with a thickness of 6.22 nm in the intrinsic $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$ region. The temperature is 300 K and the density of the carriers is $5 \times 10^{18} \text{ cm}^{-3}$. The continuous line is the result of calculations that do not take electron-electron scattering into account; the dotted line relates to calculations in which electron-electron scattering is taken into account.

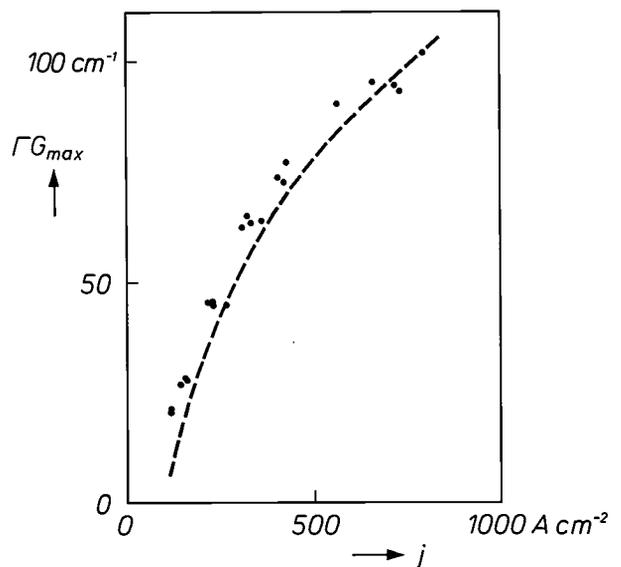


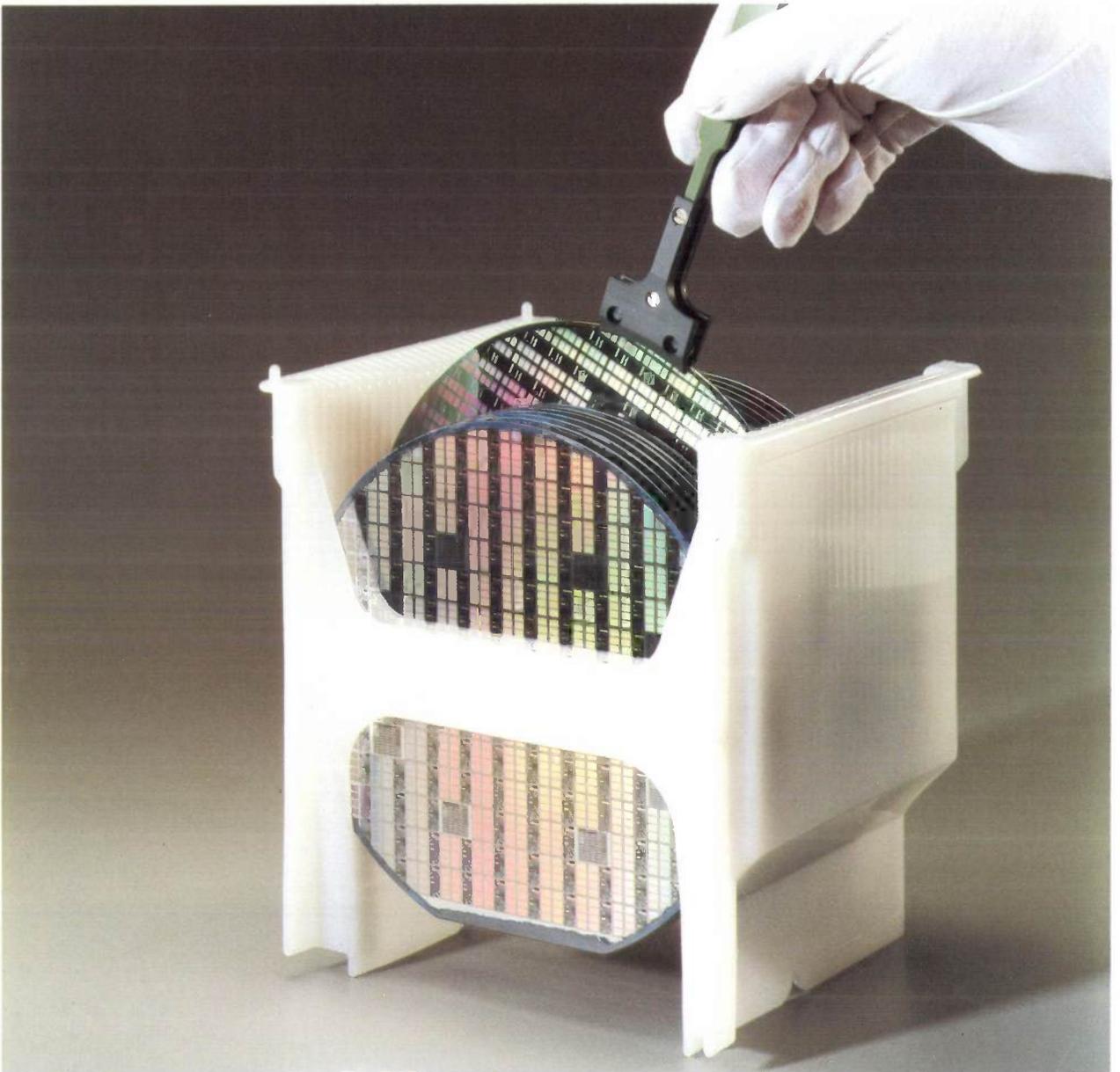
Fig. 17. Calculated (---) and measured^[8] (•) maximum gain G_{max} (multiplied by a factor Γ equal to 0.0062 for this structure and indicating the degree of light confinement) as a function of current density, i.e. the current in the quantum well per unit area, for a quantum-well laser in which there is a 2-nm GaAs layer in the intrinsic AlGaAs region. The temperature is 153 K. Electron-electron scattering is taken into account in the calculation.

well, such as electron-electron scattering, inhomogeneous current distribution in the quantum well and energy levels shifting to lower values because of the Coulomb interaction between the electrons and the holes. The dotted line represents a calculated gain spectrum in which electron-electron scattering was taken into account. The spectra calculated in this way are in reasonable qualitative agreement with measured spectra, and in some aspects there is reasonable quantitative agreement (see *fig. 17*). For complete quantitative agreement, however, all the aspects mentioned above (and others too) have to be taken into account. A fully detailed description of a GaAs/AlGaAs quantum-well laser would take us beyond the scope of this article, however.

We have attempted to provide a better understanding of the theory of the quantum well. An example is included to show that the results that can be achieved with this theory are in excellent agreement with the experimental results. We have also shown how an extended version of the theory can be used in model calculations on quantum-well lasers.

Summary. A quantum well — a thin layer of semiconductor material (GaAs) between thicker layers of a dissimilar semiconductor material (AlGaAs) — is the heart of a quantum-well laser. A theory has been developed that explains the operation of the quantum-well laser and predicts its performance. Because of the relative locations of the energy levels in GaAs and AlGaAs the carriers are confined in the GaAs layer. The quantum-mechanical theory for a particle in a box is applied to this situation and extended to include the interaction between the different bands. The theory also takes into account the creation of excitons — electrons and holes coupled in pairs. By way of illustration the theory is used for calculating the absorption spectrum of a quantum well consisting of a 7.7-nm layer of GaAs between AlGaAs layers. The agreement with the measured absorption spectrum is excellent.

^[8] S. H. Hagen, H. F. J. van 't Blik, M. J. B. Boermans and R. Eppenga, Experimental determination of the relation between modal gain and current density for AlGaAs single quantum well lasers grown by metalorganic vapor phase epitaxy, *Appl. Phys. Lett.* 52, 2015, 1988.



Submicron ICs

As announced earlier, Philips will shortly market a new integrated circuit that contains an SRAM (Static Random-Access Memory) with a capacity of 1 megabit ($= 2^{20}$ or 1 048 576 bits). This chip will be fabricated in submicron technology, i.e. the smallest details on the chip will be smaller than $1\mu\text{m}$. Other circuits are also being made in the same technology. The photograph shows a number of silicon wafers (or 'slices') with a diameter of 15 cm, packed with these circuits. These wafers were ready on the 15th April of this year. They are part of the first batch produced in the 6-in pilot

plant at the IC Centre at the Philips Research Laboratories complex in Eindhoven; the quality of the batch was high enough to test the operation of the memory circuits.

Each wafer in the photograph contains more than two hundred 256-kbit SRAMs. Since the storage of each bit requires six MOS transistors, each of the SRAMs consists of millions of components. The smallest dimension of the transistors (the gate) is only $0.7\mu\text{m}$. The coloured patterns are produced by interference of incident white light in the extremely thin films from which the circuits are formed.

Laboratory-scale manufacture of magnetic heads

J. P. M. Verbunt

Magnetic recording of information is definitely a growth activity. It is mainly used today for recording audio, video and computer data. Many successes have already been achieved, but the search continues for improvements that will reduce costs, improve quality or even lead to entirely new applications. The magnetic heads used in recording obviously have an important part to play. The article below deals with the manufacture of magnetic heads for experimental purposes, using fabrication techniques appropriate to small-scale production.

Introduction

Alexander Graham Bell's invention of the telephone in 1876 gave a great stimulus to the search for ways and means of recording sound signals on some kind of storage medium. As early as 1877 Thomas Alva Edison brought out his famous phonograph — a purely mechanical device. The first description of a system in which use was made of an electrical signal is due to a certain O. Smith and appeared on 8th September 1888 in 'The Electrical World' [1]. Smith proposed that a wire of a suitable material, containing small steel particles, could be passed through the centre of a coil. A microphone signal in the coil would then magnetize the wire in accordance with the pattern of the sound information (*fig. 1a*). The sound could be reproduced by passing the magnetized wire through the coil again and listening to the resulting induced current through headphones. The inventor only wrote about his idea, however, and never built a working device. This was done for the first time by the Dane Valdemar Poulsen, who applied for the first patent on his device in 1898 (*fig. 1b*), and followed this by an improved and more detailed American patent in 1900 (*fig. 1c*). The first major public demonstration of the device was given in the same year at the World Exhibition in Paris.

It is remarkable that Poulsen, in his American patent application, made the first explicit mention of the tele-

phone-answering machine among the potential applications. The close relation between magnetic recording and telephony also appears from the name generally used for the magnetic recording machine in the early days: the 'telegraphone'. A wide range of applications have evolved from this machine through the years, from video cassette recorders and pocket players for audio cassettes to disc and tape memories for large and small computers.

However much present-day machines may differ from one another or in comparison with the oldest versions, they all have three basic elements in common:

- the recording medium in the form of a wholly or partially magnetizable wire, tape, disc or cylinder;
- the magnetic head, which magnetizes the medium during recording ('writing') and scans it during playback ('reading');
- the transport mechanism, which provides the correct relative movements of medium and head.

Producing these three basic elements always depends on at least one of the following disciplines:

- chemistry, especially colloid chemistry;
- mechanical engineering, especially precision engineering;
- the theory of electromagnetism, especially magneto-statics and magnetodynamics.

J. P. M. Verbunt is with Philips Research Laboratories, Eindhoven.

[1] M. Camras (ed.), Magnetic tape recording (Benchmark papers in acoustics, Vol. 20), Van Nostrand Reinhold Company, New York 1985.

Nowadays electronics also plays a vital role in processing the signals before and after recording. It is this interdisciplinary aspect of magnetic recording that makes it such an interesting — if difficult — field of study.

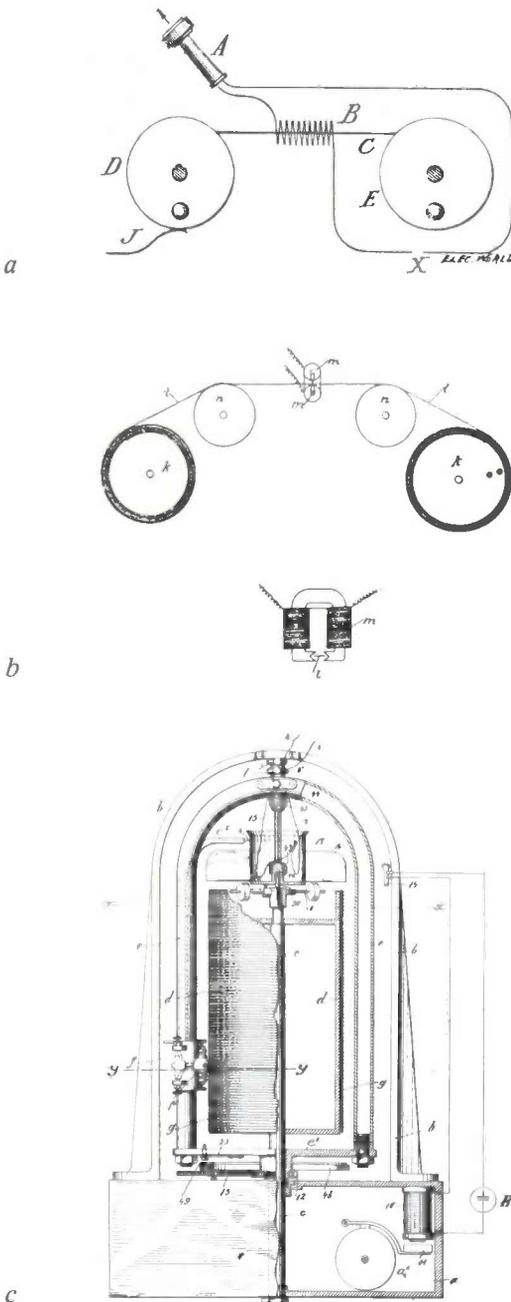


Fig. 1. *a*) Original drawing of the oldest known proposal for magnetic recording (1888). The recording medium used is a wire. A d.c. source is located at *X* during recording. The author refers to 'playback' in the following terms: '(. . .) it may be possible to insert at *X* (. . .) some intensifying apparatus, such as a battery, but which has not yet been thought out'. So here electronics literally was the missing link. *b*) Diagram from the original Danish patent taken out by V. Poulsen (1898). The recording medium here is a tape. *c*) In his first American patent (1900) Poulsen proposed the use of a stationary drum to which a helix of steel wire was fixed as the recording medium. The write and read head was mounted on a carrier shaped like an inverted U. It could move vertically along the carrier, which could rotate around the drum.

Although some of the techniques are now quite mature, research continues on ways of making better versions of the various basic elements. It is desirable, for example, to have higher information densities per unit area of the recording medium and for the recording process to be more compatible with digital rather than analog signals. This is because digital signals are mainly binary, and there are different requirements for factors such as the signal-to-noise ratio at playback.

These studies will be illustrated in this article with a description of a number of experimental magnetic heads made in the last few years. It will be useful to start with a brief recapitulation of some of the basic concepts.

Basic concepts in magnetic recording

Magnetic recording today is based on the same principles^{[2][3]} as in Poulsen's original recording machine: an electric current, representing the information to be recorded, flows in a coil (fig. 2). Inside the coil there is a core of magnetic material, usually ring-shaped. When the current in the coil is varied, a varying magnetic field is excited in the core. A discontinuity ('gap') at a defined position in the core makes the magnetic field spread outside the recording head, or 'write head', so that it can magnetize a recording medium that bridges the gap — in this case a tape. Later on, the magnetized tape can be drawn past a playback head, or 'read head', in which a varying magnetic flux is excited, and this in turn induces an electrical signal in the coil. This signal represents the original information. In principle the same head can be used for both recording and playback; for playback or 'read-out' it is also possible to use the magnetoresistance effect^{[4][5]} instead of the conventional electromagnetic-induction method. In that case a playback head of a different design is used.

If the preferred direction of magnetization on the tape is parallel to the direction of tape travel, the recording is said to be longitudinal^[6]. This has long been the conventional form of recording. Through the years the minimum wavelength λ (fig. 2*b*) has been reduced by a factor of 10^3 to 10^4 , so that a very much higher information density can now be achieved (fig. 3). To obtain even higher information densities on the tape it will probably be necessary to use 'perpendicular recording', with the preferred direction of magnetization perpendicular to the surface of the medium.

Nowadays the recording medium usually consists of a thin magnetizable layer uniformly deposited on a non-magnetic substrate.

Both the magnetizable part of the medium and the core of the magnetic head consist of magnetic mat-

erial. Their properties, however, have to be very different. This can be demonstrated by means of the $B-H$ characteristic, which shows the magnetic flux density B as a function of the magnetic field-strength H (fig. 4). Since the value of B depends not only on the instantaneous value of H , but also on the earlier history of the specimen, the $B-H$ characteristic has the shape of the familiar hysteresis loop. As H increases, so does B , first relatively steeply (until a saturation flux density B_s is reached), and then less steeply. When H falls to zero, B does not become zero but stays at a remanence value B_r ; for B to drop to zero an opposing magnetic field of strength $-H_c$ has to be applied ('c' for coercivity). In the magnetic layer of the tape we want a high value of B_r , because this gives a high magnetization after recording. We also want a high value of H_c (fig. 4a), because this gives a high degree of insensitivity to the demagnetizing effect of adjacent parts of the tape and to

Ned. Philips Bedrijven B.V.
 PHILIPS RESEARCH LABS.
 LIBRARY WY - 1
 P.O. Box 80 000
 5600 JA EINDHOVEN
 THE NETHERLANDS

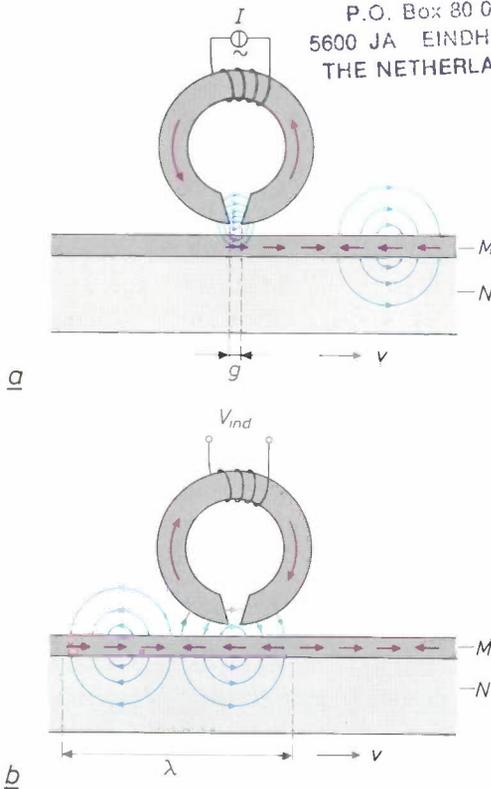


Fig. 2. Principle of magnetic recording with a ring-shaped record and playback head. a) During recording an alternating current I excites a varying magnetic field in the head. This field emerges from the magnetic circuit at the head gap. A recording medium drawn past this gap is then magnetized in a pattern determined by I . b) During reproduction the recording medium is again drawn past the head. The magnetization produces a varying magnetic field in the head and an alternating voltage V_{ind} is induced in the coil. The recorded information is recovered from V_{ind} . The direction of magnetization of the materials is indicated by red arrows; the blue arrows indicate the magnetic lines of force. M magnetizable layer, N plastic substrate, v relative velocity, g gap length, λ wavelength of the recorded magnetic pattern.

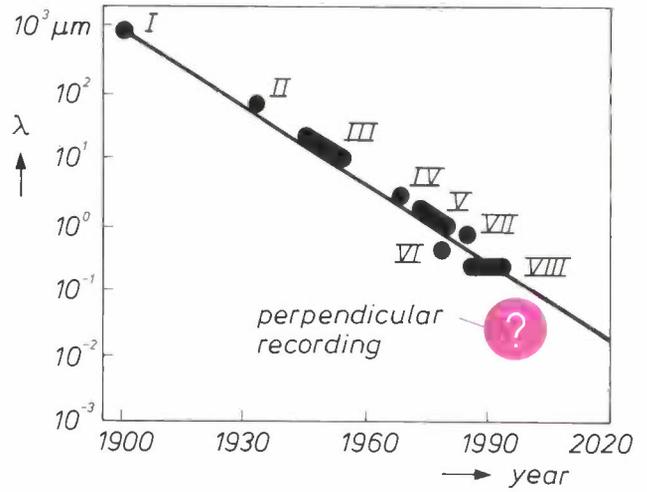


Fig. 3. As recording media are improved the minimum wavelength of the recorded magnetic pattern decreases and the information density becomes higher. So far longitudinal recording has mainly been used; one of the ways in which the information density can be increased still further is by using perpendicular recording. I Poulson's steel wire; II magnetic tape; III tape for reel-to-reel recorder; IV compact cassette tape; V video recorder tape (Beta, VHS, V2000); VI video recorder tape (8-mm metal); VII tape for digital video; VIII thin metal tape.

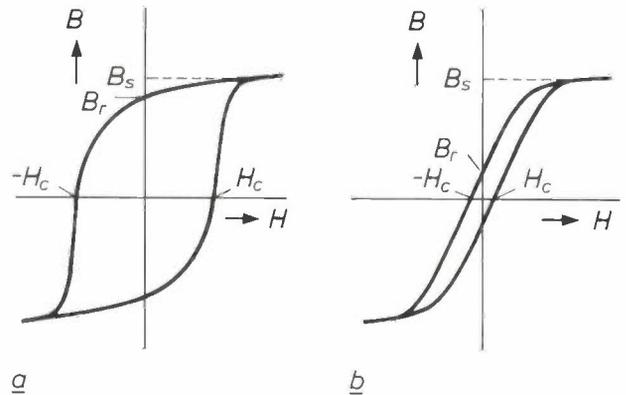


Fig. 4. Magnetic materials can be characterized by a $B-H$ characteristic or hysteresis loop, where H is the external magnetic field-strength (in A/m) and B is the magnetic flux density (in T). For magnetic recording media B_r and H_c should have high values, as in (a), and for the magnetic head material B_r should have a high value and B_r and H_c should have low values as in (b).

[2] W. K. Westmijze, The principle of the magnetic recording and reproduction of sound, Philips Tech. Rev. 15, 84-96, 1953/54.
 [3] W. K. Westmijze, Studies on magnetic recording (Thesis), published in six parts in Philips Res. Rep. 8, 148-157, 161-183, 245-255, 255-269, 343-354, 354-366, 1953. (Reprints of these articles can be found in [1].)
 [4] W. J. van Gestel, F. W. Gorter and K. E. Kuijk, Read-out of a magnetic tape by the magneto-resistance effect, Philips Tech. Rev. 37, 42-50, 1977.
 [5] M. G. J. Heijman, J. H. W. Kuntzel and G. H. J. Somers, Multiple-track magnetic heads in thin-film technology, to be published shortly in this journal.
 [6] It is not easy to make an accurate analysis of the magnetization processes both in and between the head and the recording medium. In the past greatly enlarged scale models have been a useful aid to understanding. See for example: D. L. A. Tjaden and J. Leyten, A 5000:1 scale model of the magnetic recording process, Philips Tech. Rev. 25, 319-329, 1963/64. (A reprint of this article is also given in [1].)

any interfering fields that may be present. In the material of the magnetic head we want B_s to have a high value, but B_r and H_c to have low values (fig. 4b) to ensure that the instantaneous characteristics of the magnetic head will be virtually independent of signals recorded or played back earlier.

The gap in the magnetic material of the head is of crucial importance in the magnetic circuit, since this is where the magnetic coupling takes place between the head and the medium (see fig. 2). During recording the 'fringe field' at the head gap is responsible for magnetizing the medium; during playback as much of the external magnetic field of the medium as possible should follow the magnetic circuit of the head, and as little as possible should be 'short-circuited' via the gap. This is done by using a material of high relative permeability (μ_r) for the head and making the cross-section of the head at the gap (A_g) smaller than the cross-section of the rest of the magnetic circuit (A_m).

The importance of the relative permeability μ_r of the magnetic head and of the cross-sectional ratio A_m/A_g in the recording process can easily be demonstrated. For simplicity we assume that the gap of a magnetic head contains only a uniform magnetic field H_g , and that the fringe field H_f is negligibly small compared with H_g .

We first consider a closed ring of magnetic material with a coil of N turns wound round it, with a current I in the coil (fig. 5a). The magnetic field-strength H_m in the ring is then given by

$$\oint H_m dl = NI. \tag{1}$$

If the length of the centre-line through the ring is L , we have:

$$H_m = NI/L. \tag{2}$$

If we now make a gap of length g in this ring (fig. 5b) and we reduce the cross-section at the gap from A_m to A_g , then, neglecting H_f , we find from (1):

$$g H_g + (L - g) H_m = NI. \tag{3}$$

Also, the total magnetic flux in the gap is the same as in the rest of the ring, so that

$$\Phi_g = \Phi_m \tag{4}$$

or (using the magnetic flux densities B_m and B_g)

$$B_g A_g = B_m A_m \tag{5}$$

or

$$\mu_0 H_g A_g = \mu_0 \mu_r H_m A_m. \tag{6}$$

Combining (6) and (3):

$$H_g = \frac{NI}{g + (L - g) \frac{A_g}{\mu_r A_m}}. \tag{7}$$

At given values of N , I , L and g the magnetic field-strength in the gap is highest when μ_r and the ratio A_m/A_g are as large as possible. In practical situations, where the fringe field H_f neglected above is the crucial quantity, the relation between H_f and the quantities μ_r and A_m/A_g is much the same as for H_g in equation (7).

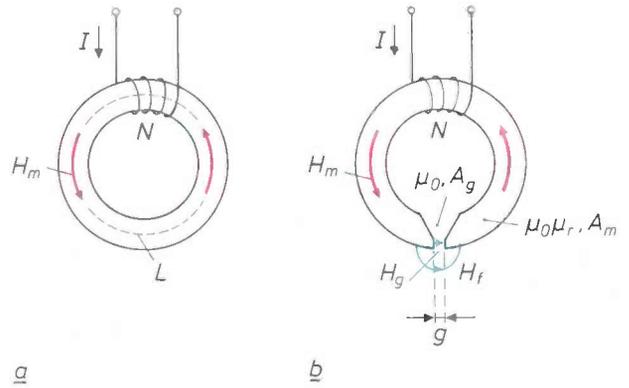


Fig. 5. a) Closed ring of magnetic material with a coil wound round it. A current I flows in the coil. H_m magnetic field-strength; L length of the centre-line through the ring; N number of turns. b) Simplified model of a ring-shaped magnetic head. g gap length; H_m , H_g , H_f magnetic field-strength; $\mu_0 \mu_r$, μ_0 magnetic permeability; A_m , A_g cross-section.

Gap dimensions

The characteristics of a magnetic head depend to a very great extent on the length g of the gap^[7](see fig. 5b). The gap length is in fact one of the most important factors determining the maximum signal frequency that can be handled in magnetic recording; in this respect the playback process is more critical than the recording process. To make this clear let us refer again to fig. 2. If we are recording a signal frequency of f Hz and the medium (a tape) is travelling at a relative velocity of v m/s with respect to the magnetic head, then a periodic magnetic pattern is recorded on the tape with a wavelength λ given by $\lambda = v/f$ m. In the recording process the magnetization of the tape after passing the head is mainly determined by the magnetic field at the end of the gap (the 'trailing edge'), rather than the field over the whole length of the gap. During the playback process, on the other hand, the gap plays a significant part over its whole length. If the wavelength of the magnetic pattern on the tape is exactly equal to the gap length, both ends of the gap will always be opposite to places on the tape with the same magnetization. There will then be no varying magnetic flux in the magnetic head and no signal voltage will be induced. The corresponding signal frequency will consequently not be detected by the playback head and the same applies to every integer multiple of this frequency. To achieve the highest possible information density (corresponding to the smallest possible λ) it is therefore necessary to minimize the gap length. It is self-evident that reducing the width of the gap can also increase the maximum information density per unit area of the tape: the smaller

the gap width (or 'track width') the more tracks can be accommodated side by side on a given area of the tape.

Types of magnetic head

The consequences of what has been said in the foregoing about the shape of magnetic heads can be seen from the diagrams in *fig. 6*. Two relatively modern magnetic heads are illustrated. Their outside dimensions are roughly the same ($0.2 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$). The head in *fig. 6a* is made from the magnetic material MnZn ferrite and is used for recording analog video signals in the V2000 system and the VHS system. The head in *fig. 6b* is also of MnZn ferrite and has not only a much smaller gap length but also a much smaller gap width, $10 \mu\text{m}$ instead of $23 \mu\text{m}$, so that a narrower magnetic track can be recorded. This head is designed for recording digital video signals. The use of a narrower track means that both the useful signal and the background-noise signal will be weaker at playback. They are not decreased in proportion, however: every time the gap width is halved there is a 3-dB reduction in the signal-to-noise ratio.

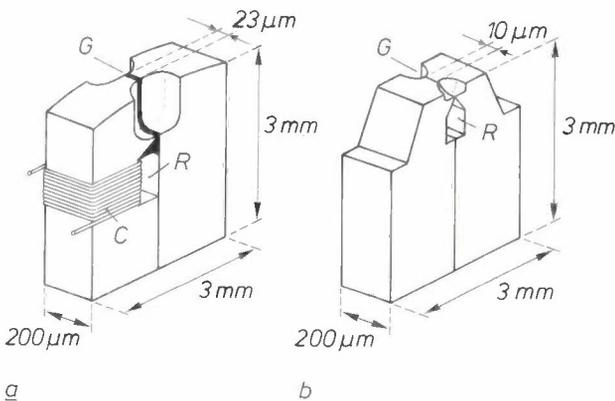


Fig. 6. Two examples of modern magnetic heads for recording video signals. *a*) Magnetic head designed for analog video signals and used in the V2000 system and the VHS system. *b*) Magnetic head designed for digital video signals. This head has a gap *G* of even smaller length and width. *R* coil chamber. (For clarity the coil *C* has been omitted.)

With reduced gap widths it is also possible to produce multiple heads, so that a number of magnetic tracks can be recorded on a medium simultaneously but independently. Techniques of this type are very suitable for recording digital signals, since it is easy to obtain a number of parallel binary signals. Applications have also been found for analog signals, however^[5]. These multiple heads are often made by photolithographic processes, as used in the manufac-

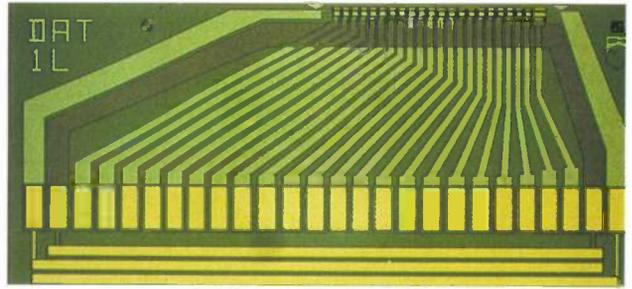


Fig. 7. Example of a 22-track magnetic head of the thin-film type. The manufacturing technology is similar to that used for producing semiconductor chips. The horizontal dimension of the head shown here is only about 5 mm.

ture of integrated circuits. They are then referred to as 'thin-film heads'. An example is shown in *fig. 7*. The tape is supposed to be travelling in the direction perpendicular to the plane of the photograph here. We shall not consider multiple heads further in this article.

One of the ways of increasing information density in magnetic recording is to increase the coercivity of the magnetic material of the recording medium. To record on a material of higher coercivity it is necessary, however, to use a head material that gives a higher saturation flux density, particularly near the gap^[8], where the cross-section of the magnetic material is smaller and the magnetic flux density *B* is higher (see small print on page 154). Philips therefore use a material such as NiFe on both sides of the gap and an Fe-Al-Si alloy (known as 'Sendust'), while the rest of the magnetic circuit is made of ferrite. The saturation flux density is then 0.8 T to 1.0 T compared with 0.5 T for MnZn-ferrite. A magnetic head made on this principle is illustrated in *fig. 8a*. Heads of this type are referred to as *metal-in-gap heads*^[9].

Another way of producing a magnetic head for high-coercivity tape is illustrated in *fig. 8b*. This is a *sandwich head*, with the thickness of the magnetic circuit nowhere more than $18 \mu\text{m}$. This magnetic layer consists of an amorphous ribbon of an FeCo alloy with a saturation flux density of 0.8 to 1.0 T. It is enclosed between two much thicker parts of non-mag-

^[7] The gap length is the dimension of the gap in the direction of relative movement of the magnetic head and the medium during recording and playback; the dimension in the direction perpendicular to the medium is called the gap height, and the dimension in the direction perpendicular to the other two is called the gap width. In the examples given here the gap width is usually greater than the gap length.

^[8] F. J. Jeffers, R. J. McClure, W. W. French and N. J. Griffith, Metal-in-gap record head, IEEE Trans. **MAG-18**, 1146-1148, 1982.

^[9] C. W. M. P. Sillen, J. J. M. Ruigrok, A. Broese van Groenou and U. Enz, Permalloy/Sendust metal-in-gap head, IEEE Trans. **MAG-24**, 1802-1804, 1988.

netic material. In the next section we shall look more closely at the manufacture of these sandwich heads.

First, however, we shall look at a magnetic head that could well have an important part to play in the future in increasing the information density through the application of perpendicular recording. In this

wheel — a process known as melt spinning. This gives a ribbon only 30 μm thick, which has an amorphous structure because of the extremely rapid quenching. The ribbon is polished with diamond paste until the required thickness of 18 μm is reached^[11]. The ribbon is then coated on both sides with a quartz layer

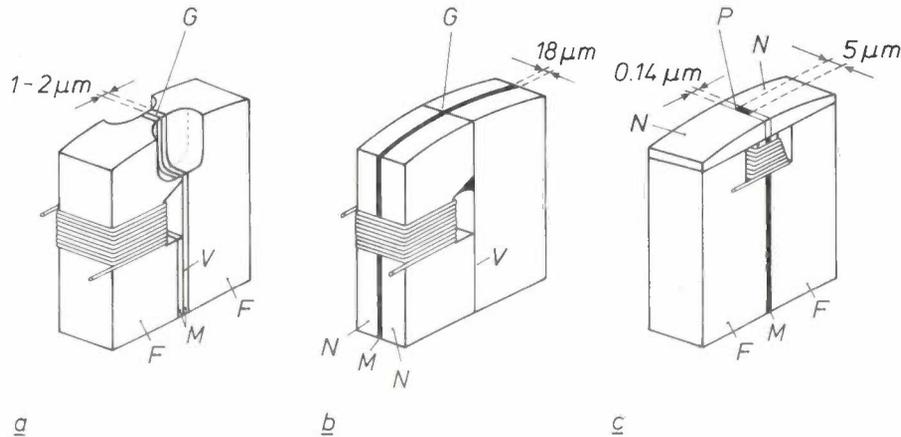


Fig. 8. Three types of magnetic head designed to give a higher information density per unit area of the recording medium. They have the same outside dimensions as the heads in fig. 6 (0.2 mm \times 3 mm \times 3 mm). *a*) Metal-in-gap head. *b*) Sandwich head. *c*) Magnetic head for perpendicular recording. The heads in (*a*) and (*b*) are designed for high-coercivity media; these heads therefore have a higher than usual saturation flux density (0.8 to 1.0 T). The head in (*c*) does not have the well-known structure of an almost completely closed magnetic circuit with a narrow gap; the magnetic circuit is shaped rather like a letter W capped by two non-magnetic pieces. Recording and playback take place here with a very thin soft-magnetic pole *P* between the two non-magnetic pieces. *F* ferrite; *G* gap; *M* magnetic material; *N* non-magnetic material; *V* gap filler material (non-magnetic).

method, as mentioned above, the recording medium is magnetized in the direction perpendicular to its surface. The magnetic head now required is no longer an almost completely closed circuit; in the example shown in fig. 8c the magnetic circuit of the head is shaped rather like a 'W' and is covered at the top with a block of non-magnetic material. A magnetic head of this type is known as a *single-pole head* and requires special types of recording media. A considerable amount of research on perpendicular recording is now under way^[10].

The manufacture of sandwich heads

To give some idea of the nature and precision of the processes and operations used in the manufacture of magnetic heads, the manufacture of a sandwich head will be described step by step. The original material consists of an amorphous metallic ribbon 18 μm thick sandwiched between two layers of ceramic material, such as barium titanate (fig. 9). The metallic ribbon is produced by spraying a liquid metal alloy, consisting mainly of cobalt and iron, on to a rapidly rotating

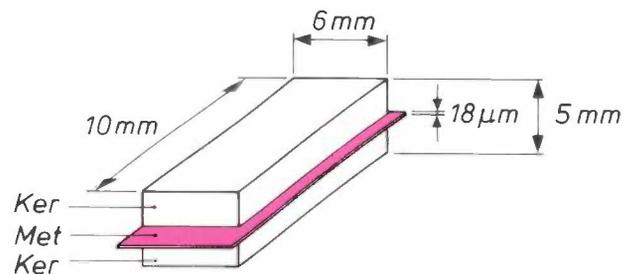


Fig. 9. The manufacture of sandwich heads is based on a material consisting of a metallic ribbon *Met*, 18 μm thick, sandwiched between two ceramic layers *Ker*.

0.15 μm thick to ensure good adhesion, and is then bonded between the two ceramic parts with an epoxy adhesive^[12]. The permitted adhesive gap must be less than 1 μm . The basic block of about 5 mm \times 6 mm \times 10 mm thus produced is then subjected to a considerable number of mechanical operations to produce two sandwich heads of dimensions 0.2 mm \times 3 mm \times 3 mm.

Various intermediate results in the process are shown schematically in fig. 10. To remove the frayed

edges of the metallic ribbon and to arrive at accurately defined outside dimensions, the block is first of all ground carefully. Two facets are ground to avoid positioning errors in later operations (fig. 10a). The block is then sawn through its centre with a saw-cut of less than 0.4 mm width (fig. 10b). The position of

It is fairly easy to saw the block now produced into two thin bars (fig. 10f); each bar will form one head. First a radius is polished on each of them (fig. 10g) on a special lapping machine with diamond as the polishing agent and a glycol/alcohol mixture as the lubricant. The surface roughness after this operation is less

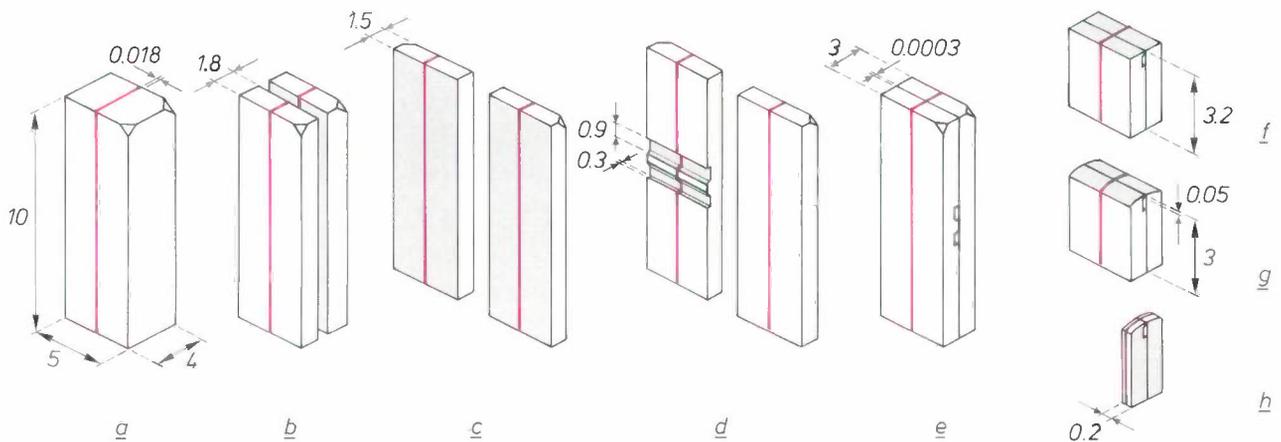


Fig. 10. Starting from the basic material in fig. 9 a sandwich head is produced after a large number of mechanical operations (all dimensions in mm). The amorphous metallic ribbon that forms the magnetic layer in the sandwich head is shown red at all stages; the surfaces of importance in the various operations are shown grey. a) Grinding and faceting; b) sawing in half; c) grinding and polishing; d) grinding the coil chamber; e) bonding; f) sawing; g) lapping to a radius; h) sawing.

the saw-cut corresponds to the place where the head gap will be. The sawing operation is also necessary to provide access to the coil-chamber location. The two halves thus produced are ground and polished very accurately to a thickness of 1.5 mm: the surface must be flat to within $0.1 \mu\text{m}$ and the roughness must be less than $0.02 \mu\text{m}$ (fig. 10c). Next, two coil chambers are produced in one of the two halves with a profiled saw (fig. 10d). The two halves can now be combined to form a single block, and this must also be done extremely accurately (fig. 10e), since the thickness of the bonding layer is exactly equal to the length of the magnetic gap of the finished head (not an air-filled gap here). The bonding layer therefore has to be $0.3 \mu\text{m}$ thick with a tolerance of the order of $0.01 \mu\text{m}$. The join can be made with an epoxy adhesive or by thermocompression with molybdenum as the adhesive layer and gold as the (non-magnetic) bonding metal. In this operation it is also necessary to ensure that the metallic ribbon aligns well in both of the constituent halves. Fig. 11a shows a plan view of an example where this is not the case; fig. 11b is well within the tolerances. The metallic ribbon (vertical; $18 \mu\text{m}$ wide) can easily be identified here, as well as the head gap (horizontal; $0.3 \mu\text{m}$ thick), which in this case consists of thin films of Mo and Au.

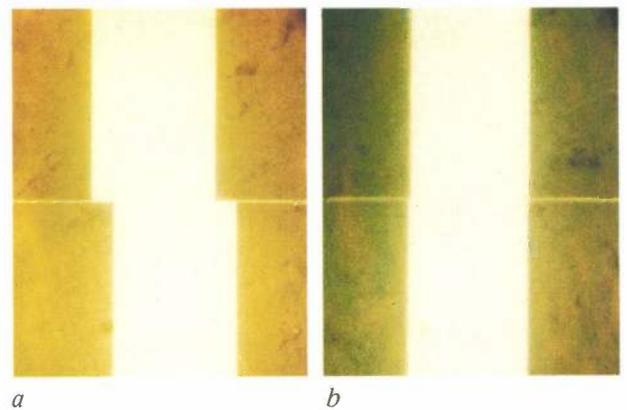


Fig. 11. The bonding operation in fig. 10e has to be extremely accurate; the gold bonding layer later forms the head gap, and its dimension ($0.3 \mu\text{m}$) is extremely critical. The metallic ribbon ($18 \mu\text{m}$ thick) in the two halves must also be aligned extremely accurately. The plan views shown here give an example (a) where the alignment has not been successful and (b) where the alignment has been successful.

[10] V. Zieren *et al.*, Efficiency improvement of one-sided probeheads for perpendicular recording on double-layer media, IEEE Trans. **MAG-23**, 2479-2481, 1987.

[11] J. G. Siekman, Polishing of brittle amorphous alloy, Wear **117**, 359-374, 1987.

[12] J. P. M. Verbunt, Lamellar magnetic core utilizing low viscosity epoxy adhesive, U.S. Patent, No. 4 713 297 (15 December 1987).

than $0.02\ \mu\text{m}$. Finally, another double sawing operation is carried out, reducing the width of the head to $200\ \mu\text{m}$ (fig. 10*h*). It is important to avoid 'crumbling' as far as possible, and this is done by using very thin diamond saws ($< 200\ \mu\text{m}$).

This completes the most critical operations, although the head is still not finished: a coil still has to

be wound through the coil chamber, the head still has to be bonded to the head-holder plate, the gap has to be given its correct height (about $35\ \mu\text{m}$) by polishing, and the head still has to be 'run in'. In the last two operations the irregularities and sharp edges remaining after the previous operations are removed. The final result can be seen in fig. 12. Fig. 12*a* shows a completely finished head, including the coils (recognizable by the two connecting wires) and the head-holder plate. Fig. 12*b* shows a detail of the coil chamber (without coil). The head gap is hardly visible in this photograph. It is filled with epoxy adhesive, which extends into the sharp corner of the coil chamber.

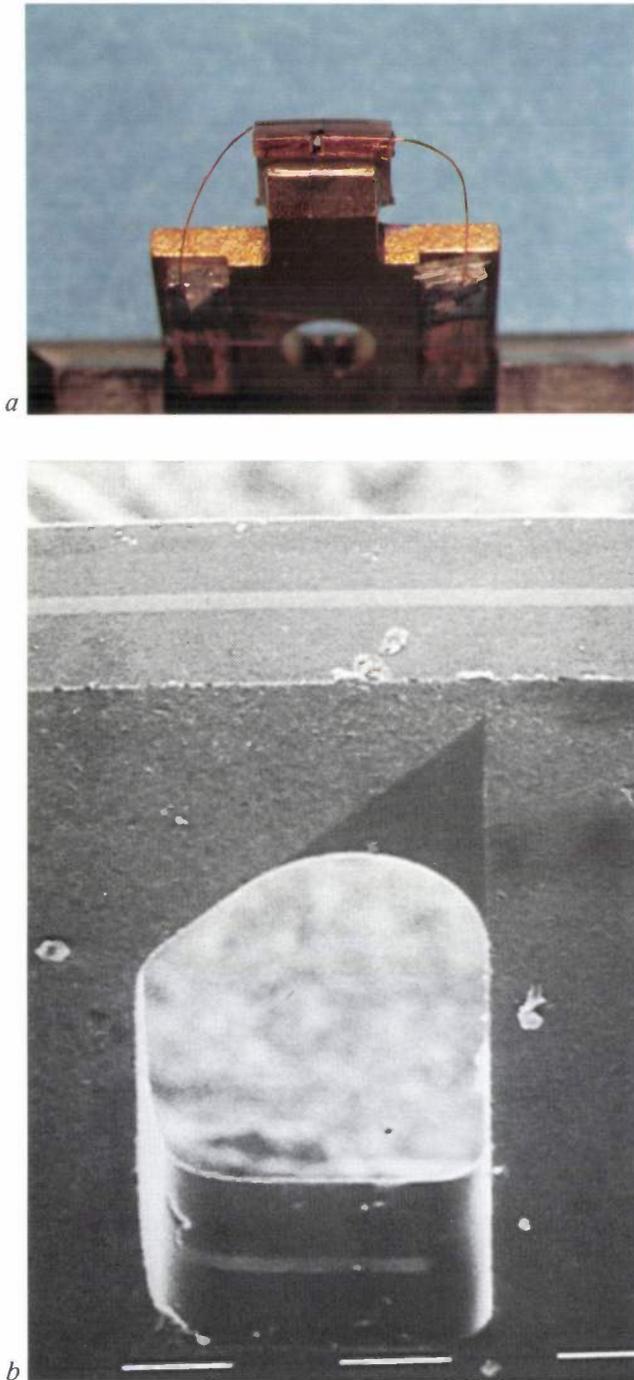


Fig. 12. *a*) Complete sandwich head including the head-holder plate it is bonded to. Since the head is viewed at an angle of about 45° , the perspective foreshortens the height. Width and height are both approximately $3\ \text{mm}$. *b*) Detail of the coil chamber of a sandwich head (without coil). The sharp corner of the hole (top right) is filled with the same epoxy adhesive as the head gap.

Laser operations

In the previous section we have seen that the manufacture of a magnetic head can require many operations. This is not true for all types of head, however, although some of the steps in the process of fig. 10 are always encountered. For long production runs, however, it is absolutely essential to simplify the machining operations. One way of doing this is to use special lasers with sufficient power at a small enough wavelength (e.g. YAG lasers). Such lasers can readily cut or drill ferrite and some of the materials used in heads for perpendicular recording such as barium titanate and zirconium oxide. Unfortunately a YAG laser cannot be used for machining glass or glass ceramic. The results of laser machining can be illustrated with the aid of fig. 13. This shows a head consisting entirely of ferrite and produced by techniques of this kind. In fig. 13*a* the head is seen from the position of the recording medium. The narrowing produced by a laser to reduce the track width can clearly be seen, as can the head gap. Fig. 13*b* shows a view seen from the coil chamber in the direction of the head gap. The ridges in the surface are due to the pulsed operation of the laser. The 'flakes' represent material that has settled on the surface after evaporation.

The fracture-gap head

We have also successfully made magnetic heads for video applications by what we call the *fracture-gap method*^[13]. The original material is a solid wafer of ferrite with the standard dimensions of a video head ($0.2\ \text{mm} \times 3\ \text{mm} \times 3\ \text{mm}$). A coil chamber is made in this wafer with a laser. Next, a thermal shock is started from the point of the coil chamber, to produce a fracture through the surface above the hole. An adhesive is then introduced into the fracture to 'fix' a magnetic gap. Using this technique we have made

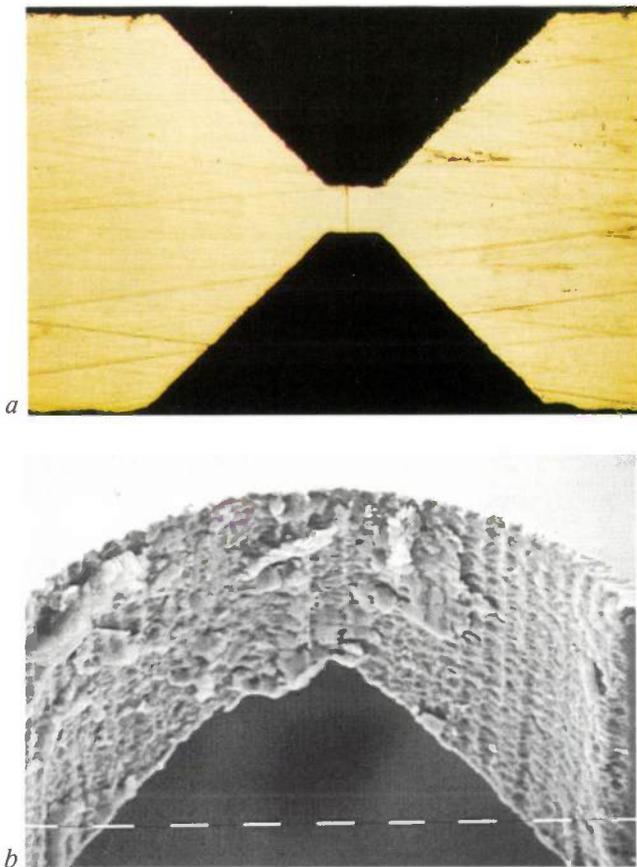


Fig. 13. Results of laser operations. *a*) The head seen from the recording medium. The photograph shows the local narrowing of the magnetic head, to reduce the track width for digital video recording. The head gap is clearly visible. *b*) View from the coil chamber looking towards the head gap (1 scale division = 1 μm).

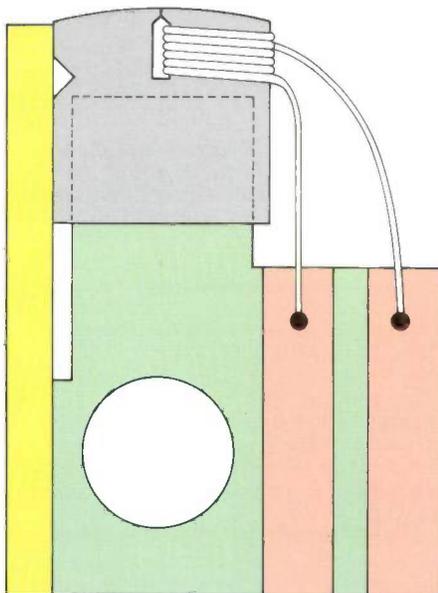


Fig. 14. A piezoelectric element (yellow) is bonded to the side of a fracture-gap head (grey) to produce a head whose gap length can easily be varied. The complete unit is mounted on a head-holder plate (green) with two conducting strips (red) for the electrical connections.

heads that give a better signal-to-noise ratio than heads produced in the conventional way. This is mainly attributable to the sharper transitions in the magnetic properties at the edges of the gap.

By keeping the gap free of adhesive and cutting a notch into the side of the head, we were able to produce a head with a variable gap length. This head had a piezoelectric strip bonded to its side (fig. 14). When a voltage is applied to this strip the notch functions as a hinge and the gap length can be varied with great precision. In fig. 15*a*, *b* and *c* gap lengths of 0, 0.3 and 0.5 μm are shown for the same head.

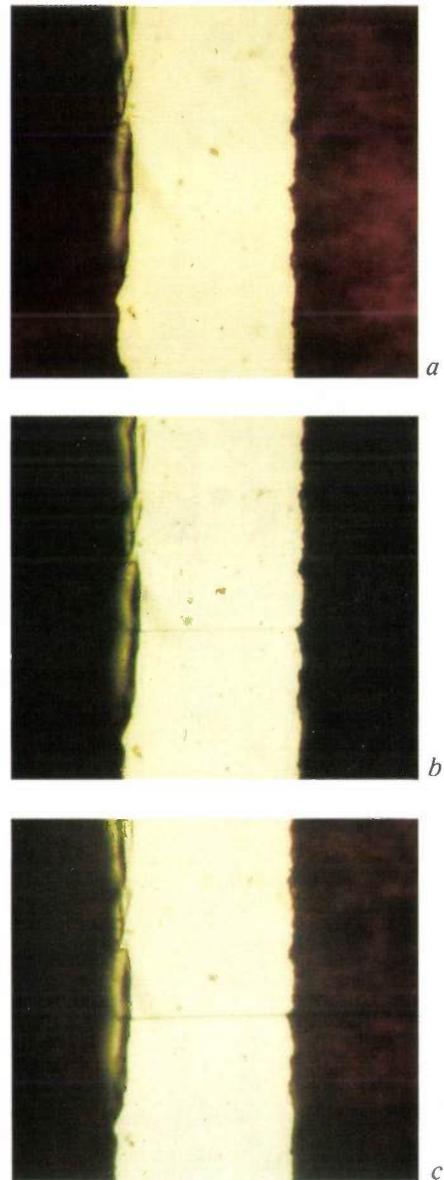


Fig. 15. Detail views of the gap of the fracture-gap head in fig. 14 with different gap lengths: *a*) 0 μm , *b*) 0.3 μm and *c*) 0.5 μm .

[13] J. P. M. Verbunt, Method of manufacturing magnetic head cores, U.S. Patent, No. 4 741 095 (3 May 1988).

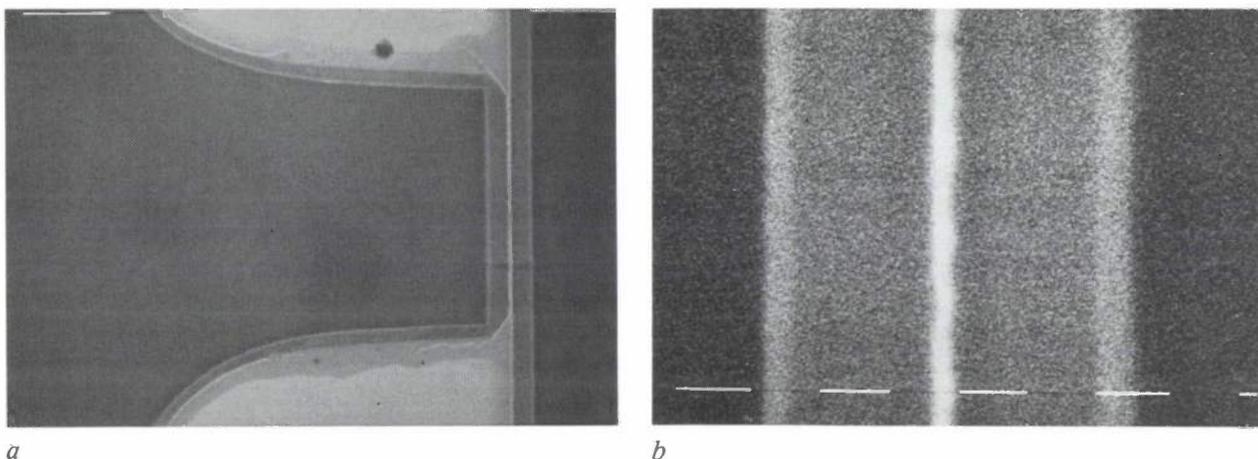


Fig. 16. In the profiled ferrite-with-glass head the magnetic circuit in one half of the head, at the gap, is significantly narrower than in the other half. *a*) View of the magnetic head at the head gap. Glass is applied on both sides of the constriction (top centre and bottom centre on the photograph). The head gap itself is filled with MoAu. On each side of the gap there is a thin film of Sendust, about 2 μm thick, separated from the ferrite by an extremely thin film of NiFe. The complete unit is thus an NiFe/Sendust metal-in-gap head. *b*) Detail of the gap (1 scale division = 1 μm).

The profiled ferrite-with-glass head

In conclusion a method will be described for producing a *profiled ferrite head*, i.e. a head whose width at the gap is smaller than in the rest of the head. If we first make a metal-in-gap head (see page 155) with a constant width over the whole head, we could still try to reduce the width at the gap by laser machining, for example. However, the ferrite reacts to the laser energy in quite a different way from the metal in and around the head gap. This metal consists of two layers of Sendust joined to the ferrite by two thin films of NiFe and separated from each other by a film of MoAu — which functions as the actual magnetic gap. Apart from the fact that the metals are more difficult to remove than the ferrite, there is also the danger of 'magnetic short-circuiting' due to fusion of the two Sendust layers. We have therefore preferred a different procedure. We give one of the two ferrite halves that form the head a profile with exactly the same width as the required gap. Next a layer of NiFe and a layer of Sendust are sputtered on to both halves of the head, and the metal to fill the gap is applied on top of these^[9]. The two halves are then joined and the open spaces (next to the actual gap) are filled from the coil chamber with a hard material. We have used a low-

melting glass for this with a coefficient of expansion almost the same as for ferrite. The presence of this material facilitates the further processing of the head and also reduces wear during subsequent use. Very good results have been achieved with this method. *Fig. 16a* shows a part of such a profiled ferrite-with-glass head around the gap; a detail of the gap can be seen in *fig. 16b*. These photographs demonstrate that this is a highly satisfactory method of manufacture.

Summary. Even a simple analysis of the principles of magnetic recording reveals a number of elementary requirements that a magnetic head has to meet. The characteristics of the magnetic circuit and the dimensions of the head gap are particularly important. In some cases the length of the head gap is only a fraction of a micron. The demand for ever-greater information densities has resulted in a wide diversity of heads (such as the metal-in-gap head, the sandwich head, the fracture-gap head and the profiled ferrite-with-glass head). All these heads are designed for longitudinal recording. Special heads have also been developed for perpendicular recording and thin-film heads for multiple recording. To give an idea of the kinds of mechanical operations encountered in the manufacture of magnetic heads, a step-by-step description is given of the laboratory-scale fabrication of a sandwich head. The use of lasers in the manufacturing process is also discussed. Some special features of the fracture-gap head and the profiled ferrite-with-glass head are described.

1938

THEN AND NOW

1988

Assembly of thermionic devices



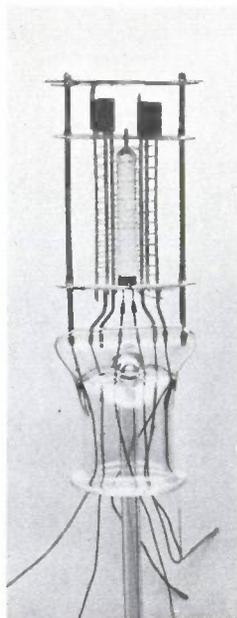
The assembly of the internal components of thermionic devices has changed a great deal in fifty years. The products are different, and so are the assembly methods. The small black-and-white photograph shows the interior of a pentode from the thirties [*]. The small colour photograph shows the electron gun of a modern Flat Square colour picture tube. In a pentode the electrons travel radially and the grids are helices of nickel wire. In the picture-tube gun the electrons travel in a fine axial beam and the grids are metal 'cans'. The beam passes through an accurately central hole in each can.

The internal components of valves used to be assembled with simple hand tools. The large black-and-white photograph shows an assembly line

in which each operator performs a simple operation on the product and passes it to her neighbour [*].

In the seventies 'work structuring' was introduced, and the responsibilities of the operators became very different. The assembly was done in small groups, with all the operators in each group taking collective responsibility for the final product. The tools used here were much more advanced.

The large colour photograph shows how the work is done in 1988. The operators are responsible for the smooth operation of the assembly process on some very expensive machinery. Each machine can deliver nearly a million guns a year for Flat Square colour tubes.



[*] From Philips Technical Review, February 1938.

A laser module for 4-Gbit/s optical communications

H. Tjassens and J. T. M. Kluitmans

Transmission rates for optical communications on glass fibre have now passed the imposing value of 1 Gbit/s. It is generally accepted that the well-known 'dual-in-line' laser package, which functions well below this almost magical bit rate, is not suitable for higher-speed applications. However, careful analysis reveals that a few simple modifications extend the operating range to at least 4 Gbit/s.

Introduction

The bit rates used in optical communications continue to increase relentlessly. Transmission systems for 565 Mbit/s are already in use, systems for 2.4 Gbit/s are under development and research has even been reported on bit rates up to 16 Gbit/s^[1]. With these increasing bit rates the laser module that performs the electro-optical conversion at the transmitter end sometimes seems to be the speed-limiting factor. However — rather surprisingly — it is not the laser diode itself, but its package that may well be the crucial element^[2]. This is also true for the dual-in-line (DIL) package now widely used in systems for bit rates up to 565 Mbit/s.

In the following we shall first describe a standard DIL laser package. Then we shall give an equivalent circuit and indicate the improvements that can easily be derived with its aid. Finally, we shall quote some improved results in the form of frequency characteristics and 'eye patterns' at 4 Gbit/s. The lasers we used in our experiments were standard commercially available 1.3- μm Philips DCPBH Fabry-Perot lasers^[3].

The DIL laser module

A standard DIL package is a rectangular block, about 1 cm \times 1 cm \times 2 cm with 14 connecting pins at the base. *Fig. 1* shows a photograph and *fig. 2* shows a schematic plan view of a DIL laser module with the cover removed. A glass fibre enters the package through one of the ends. It terminates just in front of the actual laser, which looks like a small square.

H. Tjassens and Ir J. T. M. Kluitmans are with Philips Research Laboratories, Eindhoven.

Behind the laser there is a photodiode, which monitors the light output of the laser. Next to the fibre there is a thermistor; this is included in the package to measure the temperature of the laser. The temperature can be controlled by a Peltier cooling device at the bottom of the package; all the other components are mounted on a small copper plate (the 'base carrier') fixed to the Peltier cooler. The laser itself is mounted on a smaller subcarrier that also supports an insulated stand-off. The stand-off is connected to the laser via a tiny bonding wire and to pin 9 by a rather

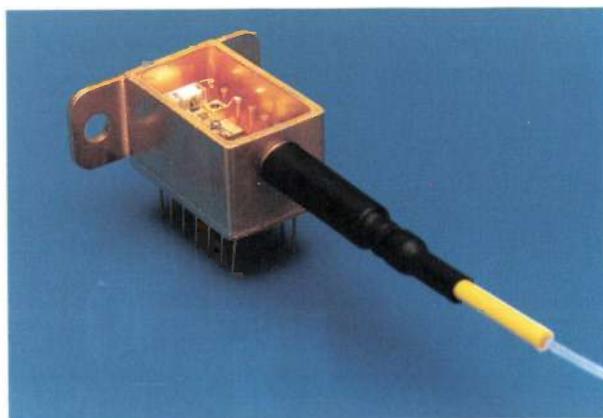


Fig. 1. Photograph of a DIL laser module with cover removed.

- [1] R. S. Tucker *et al.*, 16 Gbit/s optical time-division-multiplexed transmission system experiment, Opt. Fiber Commun. Conf., New Orleans, La., 1988, Tech. Digest, p. 149.
- [2] D. S. Alles, Tutorial on device packaging, Opt. Fiber Commun. Conf., Reno, Nev., 1987, Tutorial Sessions, pp. 237-249.
- [3] A. Valster, L. J. Meuleman, P. I. Kuindersma and T. van Dongen, Improved high frequency response of InGaAsP double channel buried hetero-structure lasers, Electron. Lett. 22, 16-18, 1986.

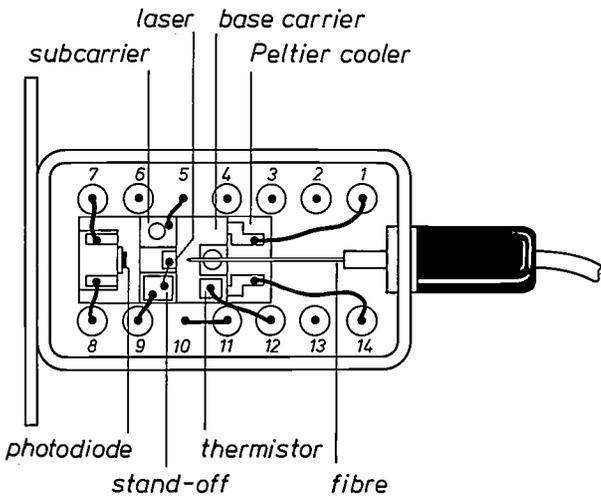


Fig. 2. Schematic top view of the module of fig. 1. The numbers 1 to 14 identify the connecting pins at the bottom.

thicker lead. Leads are also used to connect the Peltier cooler, the thermistor, the subcarrier and the photodiode to the pins of the package. All pins except 5 and 10 are insulated from the bottom of the package by a glass feed-through; pins 5 and 10 are not insulated.

The problem

Since the laser module operates at high frequencies (>1 GHz), bonding wires and leads cannot be considered as simple ideal conductors. Their inductance must now be taken into account. Similarly, the capacitance of the feed-throughs for the pins is now no longer negligible.

When these inductances and capacitances are included we can derive an equivalent circuit for the DIL laser module; this is shown in fig. 3. Table I lists the various components and their values; the values were obtained from microwave impedance measurements. The circuit of fig. 3 enables us to make a computer simulation of its electrical behaviour as a function of frequency. In the simulation the voltage between pins 5 and 9 is taken as the input quantity and the current through R_1 is taken as the output quantity. This current is directly proportional to the light output of the laser. The result of this simulation is represented by the dashed line in fig. 4; it agrees very well with the measured small-signal response (continuous line). Both curves have pronounced dips at 1.1 GHz and 1.95 GHz. These dips limit the high-frequency performance of the complete module.

A closer look at the equivalent circuit of fig. 3 shows why these dips occur: they are caused by the presence of the Peltier cooler. The dip at 1.1 GHz is due to a resonance associated with L_{i5} , L_{b5} , C_{p1} and C_{p2} . The dip at 1.95 GHz is due to a resonance associated with L_{b1} , L_{b14} , L_{i1} , L_{i14} , C_{t1} and C_{t14} .

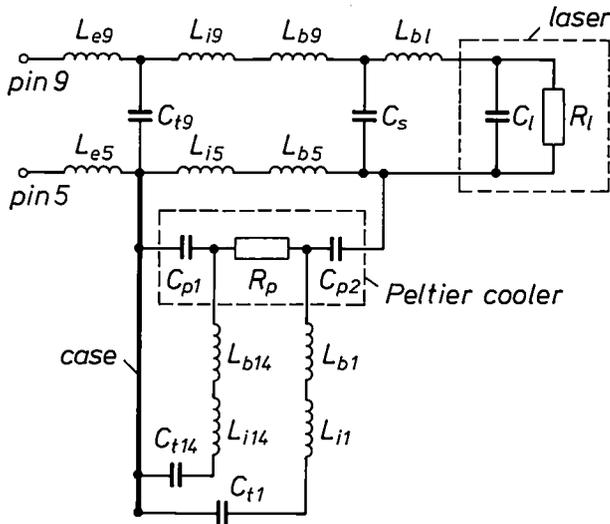


Fig. 3. Equivalent circuit for the DIL laser module. (See also Table I.)

Table I. Component values for the equivalent circuit of fig. 3.

Component	Description	Value
L_{en}	Inductance of external part of pin n ($n=5, 9$)	1 nH
L_{in}	Inductance of internal part of pin n ($n=1, 5, 9, 14$)	3 nH
L_{bn}	Inductance of internal connection to pin n $\begin{cases} (n=5, 9) \\ (n=1, 14) \end{cases}$	2 nH 10 nH
C_{tn}	Feed-through capacitance of pin n ($n=1, 9, 14$)	0.6 pF
L_{b1}	Inductance of bonding wire between stand-off and laser	0.8 nH
C_s	Stand-off capacitance	0.2 pF
C_l	Laser diode capacitance	5 pF
R_l	Laser diode resistance	6 Ω
C_{p1}, C_{p2}	Peltier cooler capacitance	7.6 pF
R_p	Peltier cooler resistance	0.3 Ω

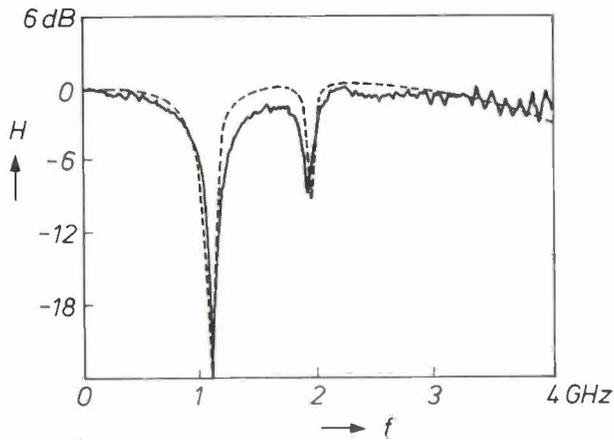


Fig. 4. Small-signal response H as a function of frequency f for the original configuration of the DIL laser module ($L_{i5} + L_{b5} = 5$ nH). Dashed line: simulation results; continuous line: measured results.

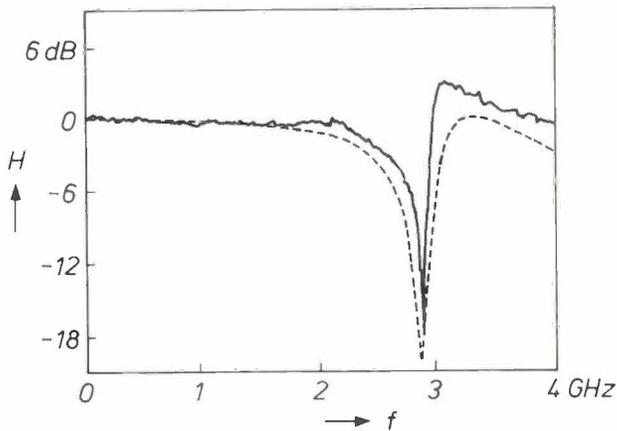


Fig. 5. Small-signal response H as a function of frequency f for the improved configuration of the DIL-laser module ($L_{i5} + L_{b5} = 0.8$ nH). Dashed line: simulation results; continuous line: measured results.

The solution

Once the problem has been defined in this way, the solution is not very difficult to find: the electrical coupling between the laser and the Peltier cooler has to be decreased. This can be done by reducing the value of the inductances L_{i5} and L_{b5} , which in turn can be done by improving the earthing (grounding) of the base carrier in the original configuration. In the improved configuration pins 5 and 10 are replaced by vertical conducting posts attached to the side walls of the package. The base carrier is also connected to the two conducting posts by bonding ribbons instead of a single-wire lead to pin 5 in the old configuration. In this way the value of the residual inductance $L_{i5} + L_{b5}$ is reduced from its original value of 5 nH to only 0.8 nH. This relatively simple change has a dra-

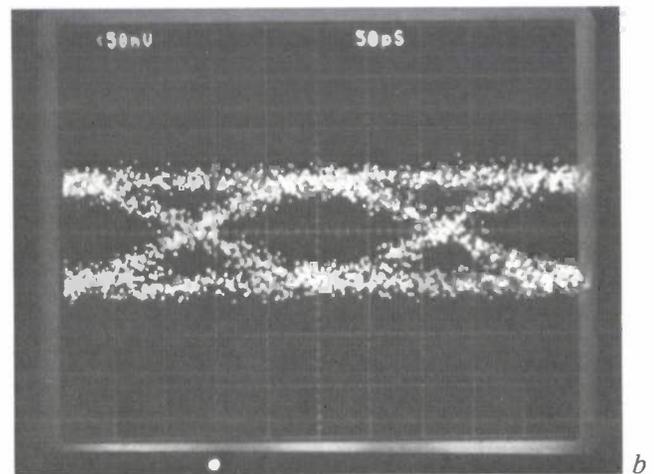
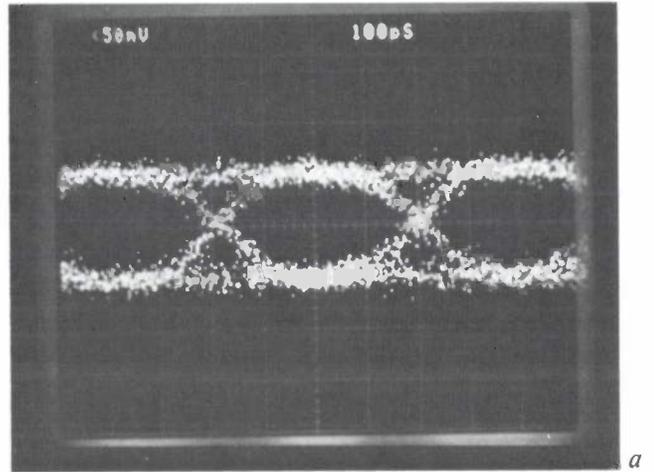


Fig. 6. Eye patterns of pseudo-random binary signals that have been transmitted by a DIL laser module with improved configuration. a) 2.5 Gbit/s; b) 4 Gbit/s.

matic effect on the high-frequency performance of the DIL laser module, as can be seen in *fig. 5*. The dashed curve gives the results of a computer simulation and the continuous curve represents measured results. The 3-dB bandwidth has been increased from 900 MHz to 2.6 GHz. As far as we know this is the highest value for this type of module reported in the literature.

Our improved module will easily handle bit rates well above 1 Gbit/s. This can be seen from the 'eye patterns' of *fig. 6*, which show pseudo-random binary signals at 2.5 Gbit/s and 4 Gbit/s.

Summary. Dual-in-line packages are frequently used for encapsulating laser diodes in optical transmission systems operating at bit rates below 1 Gbit/s. At higher rates this standard package does not perform well. Careful analysis explains the fall-off in performance and shows where simple modifications would considerably extend the useful frequency range. This is demonstrated by the performance at 2.5 Gbit/s and 4 Gbit/s. These are the highest values so far reported for this type of package.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brevannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale, P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

J. G. Kloosterboer	E	Network formation by chain crosslinking photopolymerization and its applications in electronics	Adv. Polym. Sci. 84	1-61	1988
A. J. den Boef	E	Two-wavelength scanning spot interferometer using single-frequency diode lasers	Appl. Opt. 27	306-311	1988
D. J. Gravesteyn	E	Materials developments for write-once and erasable phase-change optical recording	<i>ibid.</i>	736-738	1988
P. Hansen	H	Direct overwrite in amorphous rare-earth transition-metal alloys	Appl. Phys. Lett. 50	356-358	1987
R. J. A. A. Janssen, A. W. Kolf-schoten & G. N. A. van Veen	E	Time-of-flight study on the thermal etching of Al with Cl ₂	Appl. Phys. Lett. 52	98-100	1988
D. J. Olego, H. Baumgart & G. K. Celler (AT&T Bell Labs, Murray Hill, NJ)	N	Strains in Si-on-SiO ₂ structures formed by oxygen implantation: Raman scattering characterization	<i>ibid.</i>	483-485	1988
P. Blood, S. Colak & A. I. Kucharska	R, N	Temperature dependence of threshold current in GaAs/AlGaAs quantum well lasers	<i>ibid.</i>	599-601	1988
W. Meyer	H	Knowledge-based realtime supervision in CIM — the workcell controller	ESPRIT '86: Results and Achievements, Directorate General XIII (eds), Elsevier Science, Amsterdam	33-52	1987
P. Haaker, E. Klotz, R. Koppe & R. Linde	H	Generation of a subtraction mask for digital angiography based on the movement of the coronary arteries	Eur. Heart J. 8	57-64	1987
B. Dam, H. A. M. van Hal & C. Langereis	E	Triode-sputtered high-T _c superconducting thin films	Europhys. Lett. 5	455-460	1988
R. E. J. van de Grift, I. W. J. M. Rutten & M. van der Veen	E	An 8-bit video ADC incorporating folding and interpolation techniques	IEEE J. SC-22	944-953	1987
K. Chung	E	Discriminator-MLSE detection of a GTFM signal in the presence of fast Rayleigh fading	IEEE Trans. COM-35	1374-1376	1987
W. J. van Gils & J.-P. Boly	E	On combined symbol-and-bit error-control [4,2] codes over {0,1} ⁸ to be used in the (4,2) concept fault-tolerant computer	IEEE Trans. IT-33	911-917	1987
M. Hartmann, K. Witter, J. Reck & H. J. Tolle	H	Improvement of corrosion resistance of GdT _b Fe by metal coatings	IEEE Trans. MAG-22	943-945	1986
H. J. de Wit, C. H. M. Witmer & F. W. A. Dirne	E	Induced anisotropy of amorphous CoFeSiB and CoNbZr magnetic materials	IEEE Trans. MAG-23	2123-2127	1987

- D. J. Hermes (*Inst. Perception Res., Eindhoven*) H Measurement of pitch by subharmonic summation J. Acoust. Soc. Am. 83 257-264 1988
- H. Heitmann, M. Hartmann, S. Klahn, M. Rosenkranz, H. J. Tolle & P. Willich H Influence of preparation conditions on magnetic properties and aging behavior of rf diode sputtered GdTbCo films J. Appl. Phys. 61 3331-3333 1987
- H. Heitmann, A. M. J. Spruijt*, P. Willich & H. Wilting* (**Philips DuPont Optical Company, Eindhoven*) H Influence of nitrogen, oxygen, and water on magnetic properties of dc magnetron sputtered GdTbFe films *ibid.* 3343-3345 1987
- P. Hansen H Thermomagnetic switching in amorphous rare-earth transition metal alloys J. Appl. Phys. 62 216-230 1987
- K. W. Gerstenberg & W. Beyer (*Inst. für Grenzflächenforschung & Vakuumphys., Jülich*) H Gas evolution studies for structural characterization of hexamethyldisilazane-based a-Si:C:N:H films *ibid.* 1782-1787 1987
- S. D. Brotherton, J. R. Ayres, A. Gill, H. W. van Kesteren & F. J. A. M. Greidanus E, R Deep levels of copper in silicon *ibid.* 1826-1832 1987
- N. D. Young R A comparison between Sb and As implantation for the formation of highly conducting shallow silicon layers *ibid.* 3441-3443 1987
- K. C. Heasman*, P. Blood, A. R. Adams* (**Univ. Surrey*) & E. D. Fletcher R Pressure dependence of the nonradiative lifetime in GaAs/AlGaAs double-heterostructure lasers *ibid.* 3448-3450 1987
- J. R. Ayres, S. D. Brotherton, J. B. Clegg & A. Gill R Comparison of electrical defects in Ge⁺ and Si⁺ pre-amorphized BF₂-implanted silicon *ibid.* 3628-3632 1987
- J. Petruzzello, D. J. Olego, X. Chu* & J. P. Faurie* (**Univ. Chicago, IL*) N Transmission electron microscopy of (001) ZnTe on (001) GaAs grown by molecular-beam epitaxy J. Appl. Phys. 63 1783-1785 1988
- A. J. M. Kaizer (*Philips Consumer Electron. Div., Eindhoven*) Modeling of the nonlinear response of an electrodynamic loudspeaker by a Volterra series expansion J. Audio Eng. Soc. 35 421-433 1987
- P. J. M. van Laarhoven & T. A. C. M. Kalker E On the computation of Lauricella functions of the fourth kind J. Comput. & Appl. Math. 21 369-375 1988
- W. Tolksdorf, I. Bartels, H. Dammann & E. Pross H Growth of buried garnet channel waveguides by liquid phase epitaxy J. Cryst. Growth 84 323-325 1987
- B. J. Fitzpatrick N A review of the bulk growth of high band gap II-VI compounds J. Cryst. Growth 86 106-110 1988
- R. N. Bhargava N Materials growth and its impact on devices from wide band gap II-VI compounds *ibid.* 873-879 1988
- J. E. A. M. van den Meerakker E The electrochemistry of alkaline K₃Cr(CN)₆ solutions at p-GaAs J. Electroanal. Chem. 243 161-169 1988
- J. W. M. Jacobs & J. M. G. Rikken E Boiling effects and bubble formation at the solid-liquid interface during laser-induced metal deposition J. Electrochem. Soc. 134 2690-2696 1987
- P. J. Schoenmakers E Supercritical-fluid chromatography: open columns vs packed columns J. High Resolution Chromatogr. & Chromatogr. Commun. 11 278-282 1988
- B. Hill (*Univ. of Technol., Aachen*), K.-P. Schmidt, L. Borgmann, H. Meyer & G. Much H Design and application of magneto-optic image bars (LiSA) J. Imaging Technol. 13 15-29 1987
- L. Pareti*, F. Bolzoni*, M. Solzi* (**MASPEC, Parma*) & K. H. J. Buschow E Magnetocrystalline anisotropy in Nd_{2-x}Tb_xFe₁₄B J. Less-Common Met. 132 L5-L8 1987
- D. B. M. Klaassen, C. M. G. van Leuken & K. M. H. Maessen (*Univ. Utrecht*) E Excitation spectroscopy on lanthanide oxysulfides in the soft X-ray region J. Lumin. 40 & 41 696-697 1988
- F. R. de Boer*, Y. Huang*, Z. Zhang* (**Univ. Amsterdam*), D. B. de Mooij & K. H. J. Buschow E Magnetic and crystallographic properties of ternary rare-earth compounds of the type R₂Fe₁₄C J. Magn. & Magn. Mater. 72 167-173 1988
- P. Hansen, M. Hartmann & K. Witter H Magneto-optical properties of amorphous RE-FeCo alloys J. Magn. Soc. Jap. 11, Suppl. No. S1 257-260 1987
- W. Tolksdorf, H. Dammann & E. Pross H Magneto-optic garnet layers by liquid phase epitaxy *ibid.* 341-345 1987

- M. P. A. Vieggers, D. M. de Leeuw, C. A. H. A. Mutsaers, H. A. M. van Hal, H. C. A. Smoorenburg, J. H. T. Hengst, J. W. C. de Vries & P. C. Zalm *E* Oxygen content, microstructure, and superconductivity of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ *J. Mater. Res.* 2 743-749 1987
- G. A. C. M. Spierings & G. van Lijschoten *E* Monocrystalline MnZn ferrite dissolution in aqueous H_3PO_4 -HCl solutions *J. Mater. Sci.* 23 1060-1063 1988
- G. Harding & J. Kosanetzky *H* Status and outlook of coherent-X-ray scatter imaging *J. Opt. Soc. Am. A* 4 933-944 1987
- C. J. B. Ford*, T. J. Thornton*, R. Newbury*, M. Pepper*, H. Ahmed* (**Cavendish Lab., Cambridge*), C. T. Foxon, J. J. Harris, C. Roberts *R* The Aharonov-Bohm effect in electrostatically defined heterojunction rings *J. Phys. C* 21 L325-L331 1988
- E. A. Montie & J. C. M. Henning *E* Optically detected magnetic resonance from the DX centre in silicon-doped $\text{Al}_x\text{Ga}_{1-x}\text{As}$ *ibid.* L311-L316 1988
- L. W. Molenkamp, G. W. 't Hoof, W. A. J. A. van der Poel & C. T. Foxon *E, R* Interface related emission from an MBE-grown (Al,Ga)As heterostructure *J. Phys. Colloq.* 48, Suppl. No. 11 C5/127-C5/130 1987
- G. E. W. Bauer & T. Ando (*Univ. Tokyo*) *E* Theory of quantum well excitons in electric and magnetic fields *ibid.* C5/215-C5/218 1987
- H. H. A. Smit*, R. C. Thiel* (**Univ. Leiden*) & K. H. J. Buschow *E* On the crystal-field-induced magnetic anisotropy in B-substituted RCO_5 compounds *J. Phys. F* 18 295-306 1988
- E. H. L. Aarts, J. H. M. Korst & P. J. M. van Laarhoven *E* A quantitative analysis of the simulated annealing algorithm: a case study for the traveling salesman problem *J. Stat. Phys.* 50 187-206 1988
- M. D. Pashley, K. W. Haberern & W. Friday *N* The effect of cooling rate on the surface reconstruction of annealed silicon (111) studied by scanning tunneling microscopy and low-energy electron diffraction *J. Vac. Sci. & Technol. A* 6 488-492 1988
- J. J. Kelly, J. M. G. Rikken, J. W. M. Jacobs & A. Valster *E* Photocathodic deposition of gold alloys for Ohmic contacts to III-V materials *J. Vac. Sci. & Technol. B* 6 48-52 1988
- G. E. van Rosmalen *E* A floating-lens actuator *Jap. J. Appl. Phys.* 26, Suppl. 26-4 195-197 1987
- D. Kunz *H* Frequency-modulated radiofrequency pulses in spin-echo and stimulated-echo experiments *Magn. Resonance Med.* 4 129-136 1987
- S. Klahn, M. Hartmann, K. Witter & H. Heitmann *H* Structural relaxation and surface induced aging effects of amorphous GdTbFe films *Magnetic Properties of Amorphous Metals, A. Hernando et al. (eds), Elsevier Science, Amsterdam* 176-178 1987
- K. H. J. Buschow *E* Physical properties of ternary RE-Fe base alloys and their relationship to permanent magnet applications *Mater. Res. Soc. Symp. Proc.* 96 1-16 1987
- S. D. Brotherton, J. R. Ayres, J. B. Clegg & B. J. Goldsmith *R* Electrical characterisation of shallow pre-amorphised p⁺n junctions in silicon *Mater. Res. Soc. Symp. Proc.* 104 161-166 1988
- G. Harding, J. Kosanetzky & U. Neitzel *H* X-ray diffraction computed tomography *Med. Phys.* 14 515-525 1987
- R. E. Horstman & F. J. A. Lier *E* Digital E-beam testing *Microelectron. Eng.* 7 209-213 1987
- U. Killat *H* Asynchrone Zeitvielfachübermittlung für Breitbandnetze *Nachrichtentech. Z.* 40 572-577 1987
- A. J. Lowe*, B. C. Cavenett* (**Univ. Hull*) & M. J. Powell *R* Recombination in amorphous SiN_x -H alloys *Phil. Mag. B* 57 243-257 1988
- M. A. Hopkins*, R. J. Nicholas*, M. A. Brummell* (**Univ. Oxford*), J. J. Harris & C. T. Foxon *R* Cyclotron-resonance study of nonparabolicity and screening in $\text{GaAs}_{1-x}\text{Al}_x\text{As}$ heterojunctions *Phys. Rev. B* 36 4789-4795 1987
- G. Duggan *R* Theory of heavy-hole magnetoexcitons in GaAs -(Al,Ga)As quantum-well heterostructures *Phys. Rev. B* 37 2759-2762 1988
- P. E. Wierenga, J. A. Kubby* & J. E. Griffith* (**AT&T Bell Labs, Murray Hill, NJ*) *E* Tunneling images of biatomic steps on Si(001) *Phys. Rev. Lett.* 59 2169-2172 1987

J. Feldmann*, G. Peter*, E. O. Göbel* (* <i>Philipps-Univ., Marburg</i>), P. Dawson, K. Moore, C. Foxon & R. J. Elliot (<i>Univ. Oxford</i>)	R	Linewidth dependence of radiative exciton lifetimes in quantum wells	Phys. Rev. Lett. 59	2337-2340	1987
D. M. de Leeuw, C. A. H. A. Mutsaers, C. Langereis, H. C. A. Smoorenburg & P. J. Rommers	E	Compounds and phase compatibilities in the system Y_2O_3 -BaO-CuO at 950 °C	Physica C 152	39-49	1988
H. Ney	H	Dynamic programming speech recognition using a context-free grammar	Proc. ICASSP, Dallas, TX, 1987	69-72	1987
H. Ney, D. Mergel, A. Noll & A. Paeseler	H	A data-driven organization of the dynamic programming beam search for continuous speech recognition	<i>ibid.</i>	833-836	1987
D. Mergel & A. Paeseler	H	Construction of language models for spoken database queries	<i>ibid.</i>	844-847	1987
A. Noll & H. Ney	H	Training of phoneme models in a sentence recognition system	<i>ibid.</i>	1277-1280	1987
J. W. Slotboom, G. Streutker, G. J. T. Davids & P. B. Hartog	E	Surface impact ionization in silicon devices	Proc. IEDM '87, Washington D.C., 1987	494-497	1987
D. C. McArthur	N	A non-destructive technique for the evaluation of electric field regions in semiconductor devices	Proc. IEDM '87, Washington D.C., 1987	682-685	1987
U. Rothgordt	H	Physical imaging systems	Proc. Int. Congr. of Photographic science, Cologne 1986	17-22	1986
P. J. Severin	E	Progress in passive fibre components made using the fused-head-end method	Proc. SPIE 734	144-153	1987
S. D. Brotherton, J. R. Ayres, B. J. Goldsmith & A. Gill	R	Deep level generation centres in low temperature annealed pre-amorphised silicon	Proc. SPIE 797	26-32	1987
F. Roozeboom, A. Sikkema & L. W. Molenkamp	E	GaAs-AlGaAs multiple quantum well structures and high-power lasers grown by metalorganic vapor phase epitaxy (MOVPE) in a chimney reactor	Proc. SPIE 800	27-33	1987
A. A. Turnbull* & M. E. Cooke* (* <i>Mullard, Southampton</i>)		High resolution 64-element pyroelectric linear array IR detector	Proc. SPIE 807	92-97	1987
F. E. J. Kruseman Aretz	E	On deriving a LISP program from its specification	Sci. Comput. Program. 10	19-32	1988
J. J. Harris, D. E. Lacklison, C. T. Foxon, F. M. Selten (<i>Univ. of Technol., Eindhoven</i>), A. M. Suckling*, R. J. Nicholas* (* <i>Univ. Oxford</i>) & K. W. J. Barnham (<i>Imperial College, London</i>)	R	Wavelength-dependent photoconduction effects on the second sub-band occupancy in (Al,Ga)As/GaAs heterojunctions	Semicond. Sci. Technol. 2	783-789	1987
J. W. C. de Vries	E	Interface scattering in triple layered polycrystalline thin Au/X/Au films (X = Fe, Co, Ni)	Solid State Commun. 65	201-204	1988
S. J. Battersby, F. M. Selten, J. J. Harris & C. T. Foxon	R	Magnetoresistance effect in AlGaAs/GaAs two-dimensional electron gas structures at room temperature	Solid-State Electron. 31	1083-1088	1988
P. W. J. M. Boumans, J. A. Tielrooy* & F. J. M. J. Maessen* (* <i>Univ. Amsterdam</i>)	E	Mutual spectral interferences of rare earth elements in inductively coupled plasma atomic emission spectrometry. I. Rational line selection and correction procedure	Spectrochim. Acta 43B	173-199	1988
H. van Houten, C. W. J. Beenakker, B. J. van Wees* & J. E. Mooij* (* <i>Univ. of Technol., Delft</i>)	E	Boundary scattering modified one-dimensional weak localization in submicron GaAs/AlGaAs heterostructures	Surf. Sci. 196	144-149	1988
R. Woltjer	E	Experiments and theory on the quantum Hall effect with inhomogeneous electron density	Surf. Sci. 196	188-195	1988
M. J. M. Pelgrom	E	Delay lines with surface channel charge-coupled devices	Thesis, Enschede	1-250	1988
S. M. Wolfrum & K. W. de Graaf	E	Data acquisition system for the Sedigraph 5000D	Trends Anal. Chem. 7	13-16	1988



R. Eppenga and M. F. H. Schuurmans, Theory of the GaAs/AlGaAs-quantum well,

PhilipsTech.Rev.44, No.5, 137-149, Nov. 1988.

A quantum well — a thin layer of semiconductor material (GaAs) between thicker layers of a dissimilar semiconductor material (AlGaAs) — is the heart of a quantum-well laser. A theory has been developed that explains the operation of the quantum-well laser and predicts its performance. Because of the relative locations of the energy levels in GaAs and AlGaAs the carriers are confined in the GaAs layer. The quantum-mechanical theory for a particle in a box is applied to this situation and extended to include the interaction between the different bands. The theory also takes into account the creation of excitons — electrons and holes coupled in pairs. By way of illustration the theory is used for calculating the absorption spectrum of a quantum well consisting of a 7.7-nm layer of GaAs between AlGaAs layers. The agreement with the measured absorption spectrum is excellent.

H. Tjassens and J. T. M. Kluitmans, A laser module for 4-Gbit/s optical communications,

PhilipsTech.Rev.44, No.5, 162-164, Nov. 1988.

Dual-in-line packages are frequently used for encapsulating laser diodes in optical transmission systems operating at bit rates below 1 Gbit/s. At higher rates this standard package does not perform well. Careful analysis explains the fall-off in performance and shows where simple modifications would considerably extend the useful frequency range. This is demonstrated by the performance at 2.5 Gbit/s and 4 Gbit/s. These are the highest values so far reported for this type of package.

J. P. M. Verbunt, Laboratory-scale manufacture of magnetic heads,

PhilipsTech.Rev.44, No.5, 151-160, Nov. 1988.

Even a simple analysis of the principles of magnetic recording reveals a number of elementary requirements that a magnetic head has to meet. The characteristics of the magnetic circuit and the dimensions of the head gap are particularly important. In some cases the length of the head gap is only a fraction of a micron. The demand for ever-greater information densities has resulted in a wide diversity of heads (such as the metal-in-gap head, the sandwich head, the fracture-gap head and the profiled ferrite-with-glass head). All these heads are designed for longitudinal recording. Special heads have also been developed for perpendicular recording and thin-film heads for multiple recording. To give an idea of the kinds of mechanical operations encountered in the manufacture of magnetic heads, a step-by-step description is given of the laboratory-scale fabrication of a sandwich head. The use of lasers in the manufacturing process is also discussed. Some special features of the fracture-gap head and the profiled ferrite-with-glass head are described.

I wish to subscribe to

PHILIPS TECHNICAL REVIEW

(date)

(signature)

Please tick the appropriate box

Regular subscription 80 guilders or U.S. \$ 35.00 per volume

The subscription includes postage and will start with Vol. 44, No. 1.
Please pay when you receive our invoice.

Student's subscription 32 guilders or U.S. \$ 14.00 per volume

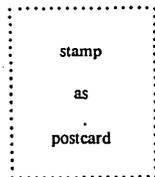
Please send a copy of your student's card or other written proof that you are a student; valid for two volumes.

Name

Initials

Title

Address



**Administration Department
Philips Technical Review**

**Philips Research Laboratories
Building WY 136**

P.O. Box 80 000

**5600 JA Eindhoven
The Netherlands**

PHILIPS TECHNICAL REVIEW

Volume 44, No. 6, December 1988

Ned. Philips Bedrijven B.V.
PHILIPS NATUURKUNDIG LAB.
BIBLIOTHEEK WY - 1
P.O. Box 80.000
5600 JA EINDHOVEN
NETHERLAND

Contents

Multi-track magnetic heads in thin-film technology	169
M. G. J. Heijman, J. H. W. Kuntzel and G. H. J. Somers <i>IC manufacturing techniques are used for fabricating 32-track playback and recording heads on a ferrite substrate</i>	
Then and Now (1938-1988)	179
Applied research — the source of innovation in consumer electronics	180
S. van Houten <i>Speech given in Nanjing in the People's Republic of China on 1st June 1988</i>	
Sound radiation from a vibrating membrane	190
J. H. Streng <i>Classical analytical mathematics and new computer programs give electroacoustic transducers with interesting characteristics</i>	
Scientific publications	200



PHILIPS

Multi-track magnetic heads in thin-film technology

M. G. J. Heijman, J. H. W. Kuntzel and G. H. J. Somers

Magnetic conductors, electrical conductors and insulators are the principal materials used in magnetic heads for sound recording. These materials can be deposited on a substrate by techniques much like those used in IC technology. This makes it possible to combine a large number of magnetic heads, each with an extremely small gap of length $1\mu\text{m}$ or less. The multi-track magnetic heads made in this way for recording 32 signals simultaneously work in the usual way with a coil. The heads that read out all these signals simultaneously do not have a coil, but an element based on the magnetoresistance effect. These multi-track magnetic heads have been developed for Philips CLS equipment (CLS stands for Communication-Logging System).

Introduction

Philips CLS equipment (CLS stands for Communication-Logging System) is designed for recording spoken messages sent by the police, fire-fighting services, banks, health-care services, and in aviation or road-traffic control. A CLS logger has to be highly reliable and capable of simultaneously recording large numbers of different speech signals containing frequencies from 300 to 3400 Hz. Tape utilization should also be very small. The latest generation of logging equipment, the CLS 8000 series, see *fig. 1*, uses magnetic-tape cassettes derived from video cassettes. Tape speed is very low, 6.6 mm/s, so that with automatic reversal the tape will run unattended for 24 hours. If the logger only records during preset times or when a

speech signal is actually presented, the period of unattended operation can be very much longer.

In a CLS logger the magnetic heads that record the signals on the magnetic tape — the recording or 'write' heads — and the magnetic heads that read the signals from the tape — the playback or 'read' heads — are different. These recording and playback heads have to meet some difficult requirements. First, they must have a long life, longer than 50 000 hours. Secondly, there should be multiple channels, so that large numbers of signals can be recorded simultaneously on a magnetic tape only 12.65 mm wide. A third requirement is that the gap in the heads should be so small that signals at a frequency of 3400 Hz can be processed at the low tape speed.

The latest equipment in the CLS 8000 series, if designed for automatic reversal of tape travel, has 32 input channels and a running time of 24 hours. Without

Dr M. G. J. Heijman and G. H. J. Somers are with Philips Research Laboratories, Eindhoven, Ing. J. H. W. Kuntzel, formerly with Philips Research Laboratories, is now with the Philips Elcoma Product Division, Eindhoven.

automatic reversal there are a maximum of 64 channels and a running time of 12 hours. The loggers have 32-track recording and playback heads. In the version with 32 channels the 32 tracks recorded in the reverse direction are 'interleaved' with the 32 tracks recorded in the forward direction. In the version with 64 channels, 64 signals are simultaneously recorded on the tape by two 32-track recording heads.

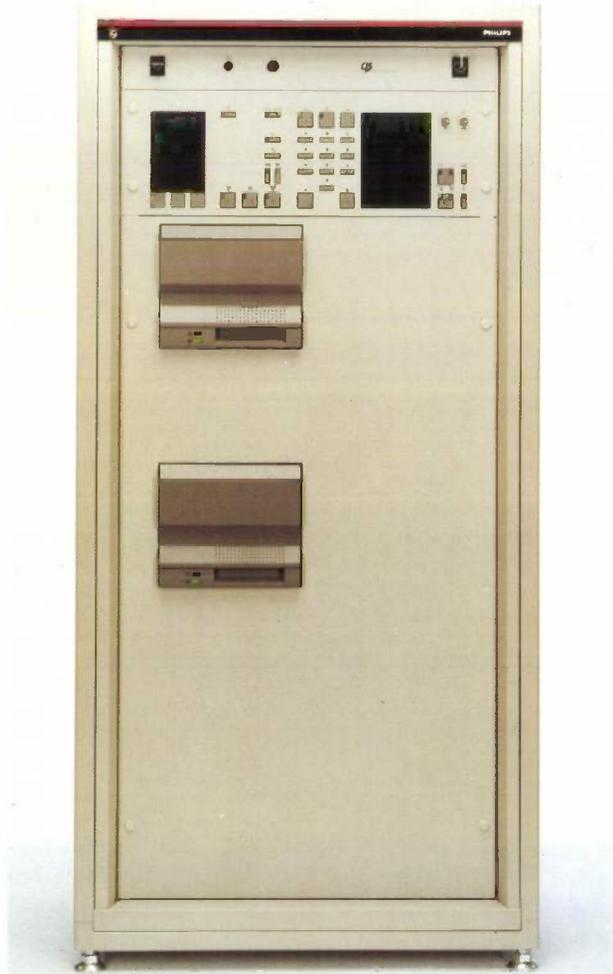


Fig. 1. One of the communication loggers in the Philips CLS 8000 series (type number is LDB 8522/32). The equipment has 32 input channels and will operate continuously, unattended, for 24 hours, or in special cases for even longer. It has a dual transport mechanism, so that two cassettes can be fitted. When one of the cassettes is full, recording is automatically switched to the other one.

The pitch of the tracks on the tape is $12.65/64 \approx 0.19$ mm. The pitch of the individual heads in a 32-track head is 0.38 mm. On automatic reversal a special mechanism moves the recording and playback heads through a distance of 0.19 mm. The effective width of a single recording head is 0.16 mm. This is also the width of a track on the tape. The tracks have to have

free space between them because they are not recorded in a perfect straight line on the tape. For the same reason the width of a playback head is slightly less than the track width: it is 0.10 mm.

Just how large is the smallest wavelength in the recorded magnetization pattern (the track)? To answer this question we must first look at *fig. 2*, which shows the principle used in recording and playback of signals by conventional magnetic heads with a coil. The wavelength λ in the magnetization pattern is equal to the ratio of the tape speed v and the frequency f . In our case the wavelength λ of a sinusoidal signal at 3400 Hz, the highest frequency that can occur, is 1.9 μ m.

The length l of the 'air gap' or 'head gap' in the magnetic circuit of the playback heads should preferably be no greater than half the wavelength λ [1]. (In magnetic heads the direction of tape travel is usually called the longitudinal direction.) The value of l adopted for CLS playback heads is 0.7 μ m, which is a fairly low value for magnetic heads. Until recently the CLS magnetic heads were made by mechanical methods. This meant that only 12 heads could be combined, so that no more than 24 signals could be recorded on a tape. The small gap length and the demand for more tracks on the tape made it necessary to adopt a more sophisticated method of manufacture: thin-film technology.

The manufacture of magnetic heads by thin-film technology usually starts with a substrate of ferrite, since this material can form part of the magnetic circuit. The magnetic 'yoke', the coil and other parts of the heads are 'planar' in the form of thin layers on the substrate. In actual use the substrate is mounted with its upper surface perpendicular to the direction of tape travel.

The first fabrication step consists in depositing a thin film by an electrochemical technique or by 'sputtering'. A photosensitive resist (a photoresist) is then applied to the film and locally illuminated through a mask. With a 'positive' resist the exposed areas are dissolved in a developer, and with a negative resist the unexposed parts are dissolved. The areas where there is no resist are then etched away. After removal of the remaining resist the next layer is deposited and subjected to the same treatment, and so on.

The processes are very like those used in IC technology. A difference is that higher temperatures are used in IC technology, because of the need for thermal diffusions. High temperatures are not required in the fabrication of magnetic heads, and indeed are undesirable since they can adversely affect the magnetic properties. The main materials used for magnetic heads are:

- permalloy (nickel-iron: 80% Ni, 20% Fe), for the magnetic circuit,
- gold, for the electrical circuit and
- silicon dioxide, for electrical insulation and for the 'air gap'.

The magnitude of the magnetic flux in the gap of a recording head, see fig. 2, is proportional to the current I in the coil and the number of turns. For CLS re-

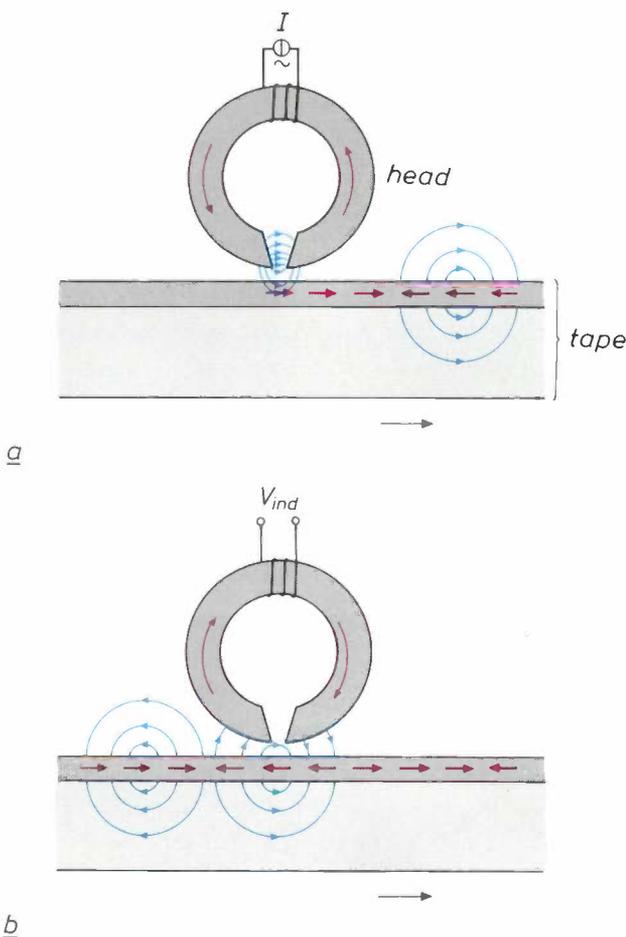


Fig. 2. Principle of recording signals on magnetic tape and playing them back in the conventional way by using magnetic heads with a coil. *a*) During recording the signal current I generates a magnetic fringing field in the head gap. This field records the signal on the tape as a magnetization parallel to the surface (red arrows). *b*) At playback the varying magnetic flux in the head, due to the fringing field around the tape (blue arrows), induces a voltage V_{ind} in the coil.

cording heads two turns per coil are sufficient, since the current can be large enough. However, an r.f. biasing field is necessary to prevent nonlinear effects due to the hysteresis of the ferromagnetic material of the tape. This bias field generates additional heat, however. As we shall see later, this problem was solved by using a pulsed high-frequency bias current instead of a sinusoidal current.

The voltage V_{ind} induced in the coil of the playback head in fig. 2 is equal to the product of the number of turns and the time derivative of the flux. The derivative is fairly small because of the low signal frequencies. To obtain a sufficiently large V_{ind} it would be necessary to have a very large number of turns for the playback head; say a hundred or more. It is not very easy to make coils with so many turns in thin-film technology. The CLS playback heads are therefore made with an element whose operation depends on the magnetoresistance effect. When this effect is used the magnitude of the signal in the playback head is essentially independent of the signal frequency.

In the rest of the article we shall first show how the magnetoresistance effect is applied. Next we shall deal with the excitation of the r.f. bias in the recording heads. Finally, we shall discuss the method of manufacture, touching on the processes, on the methods of testing and the assembly of a 32-track head.

The magnetoresistance effect

When an electric current flows in a ferromagnetic material the magnitude of the electrical resistance depends on the angle between the direction of the current and that of the magnetization^[2]. It is rather surprising to find that the magnetoresistance effect was discovered by Lord Kelvin well over a century ago, in about 1856. He wrote: 'I concluded with confidence that the electric conductivity of magnetized iron is greater across than along the lines of magnetization.' Fig. 3 shows the complicated arrangement of plates of copper and soft iron he used for the experiment he described^[3].

The operation of thin-film playback heads depends on the magnetoresistance effect in a permalloy strip. When the external magnetic field is zero, the preferred direction of magnetization will be in the longitudinal direction, in our case the x -axis. When an external magnetic field is applied in the transverse direction, along the y -axis, the resistance encountered by an electric current in the strip changes, since the angle between the current and the magnetization has changed. The curve of the change in resistance as a function of the field-strength in the y -direction, H_y , approximates to a parabola, as can be seen from the dashed curve in fig. 4a. Near $H_y = 0$, however, an approximately linear curve is desirable.

[1] J. P. M. Verbunt, Laboratory-scale manufacture of magnetic heads, Philips Tech. Rev. 44, 151-160, 1988.

[2] W. J. van Gestel, F. W. Gorter and K. E. Kuijk, Read-out of a magnetic tape by the magnetoresistance effect, Philips Tech. Rev. 37, 42-50, 1977.

[3] W. Thomson, Mathematical and physical papers, part II, Cambridge University Press, Cambridge 1884, page 307 to 327.

The change in resistance can be made linear by applying a 'barber pole' structure of electrically conducting material to the permalloy strip; see fig. 4*b*. The material used for this structure is gold. Because of the high conductivity of these areas on the strip, the angle between the current and the magnetization is 45° for $H_y = 0$. The continuous curve in fig. 4*a* shows the relation between the change in resistance and the magnitude of H_y .

Fig. 4*c* shows very schematically how the magneto-resistive element, *MRE*, is included in the magnetic

circuit of the playback head. The lines of force that come from the 'yoke' in the second air gap supply a component H_y of the magnetization in *MRE*. The change in the resistance of the strip and hence in the magnitude of the signal produced by the playback head is proportional to the magnetic flux in the yoke, and not to its derivative, as in a playback head with a coil. The amplitude of the signal in the playback head in fig. 4*c* is therefore independent of its frequency.

The form of the magneto-resistance element

As we have seen, the magneto-resistance element is built up from a ferromagnetic strip with a conducting structure on top of it; see fig. 4*b*. In general, such a

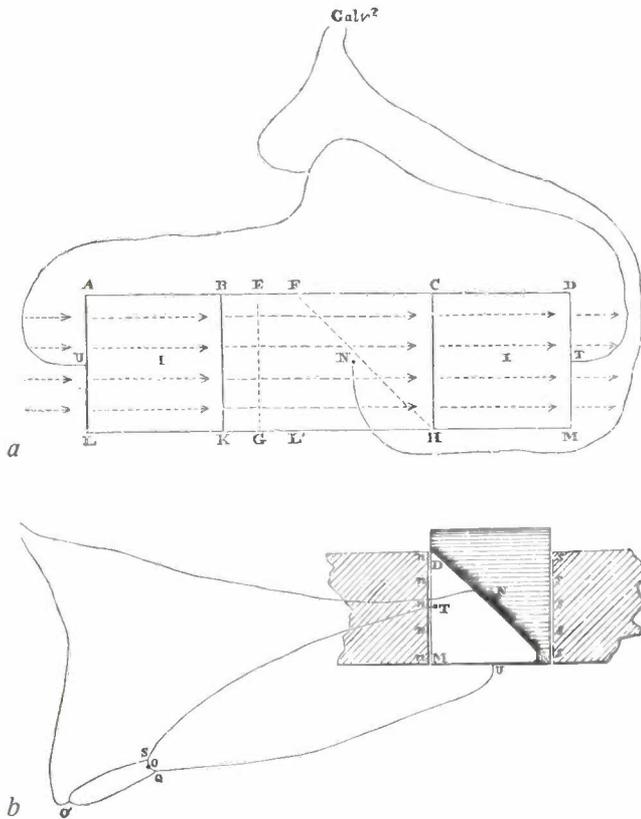


Fig. 3. The experimental arrangement that William Thomson, later Lord Kelvin, used for the first observation of the magnetoresistance effect [3]. *a*) *BCHK* copper plate. *ABKL* and *CDMH* soft-iron plates. The plates are soldered together along the lines *BK* and *CH*. The entire assembly is bent through 180° along *FH*, *CH* and *EG*, and is provided with copper connecting wires. Pieces of cardboard prevent electrical contact between the surfaces folded over one another. *b*) The assembly is now placed between the poles of a 'Ruhmkorff' electromagnet'. The poles are denoted by *n* and *s*. The two connecting wires are connected to the poles of a multiple Daniell cell, a precursor of the modern battery. The directions of current flow in the various plates are indicated by arrows in (*a*). When the electromagnet is energized, the direction of the electric current in the soft iron of *CDMH* is parallel with the direction of the magnetization; in *ABKL* they are at right angles to each other. In the bridge circuit shown here the terminals of a galvanometer are connected to the points *N* and *O*. When the electromagnet is not energized, the deflection of the galvanometer can be set exactly to zero by moving the contact point *O'* along the wire loop. Thomson discovered that the galvanometer deflected when the electromagnet was energized. The different directions of the magnetization with respect to the current in *CDMH* and *ABKL* apparently unbalance the bridge, and thus cause a difference in the electrical resistance of the soft iron.

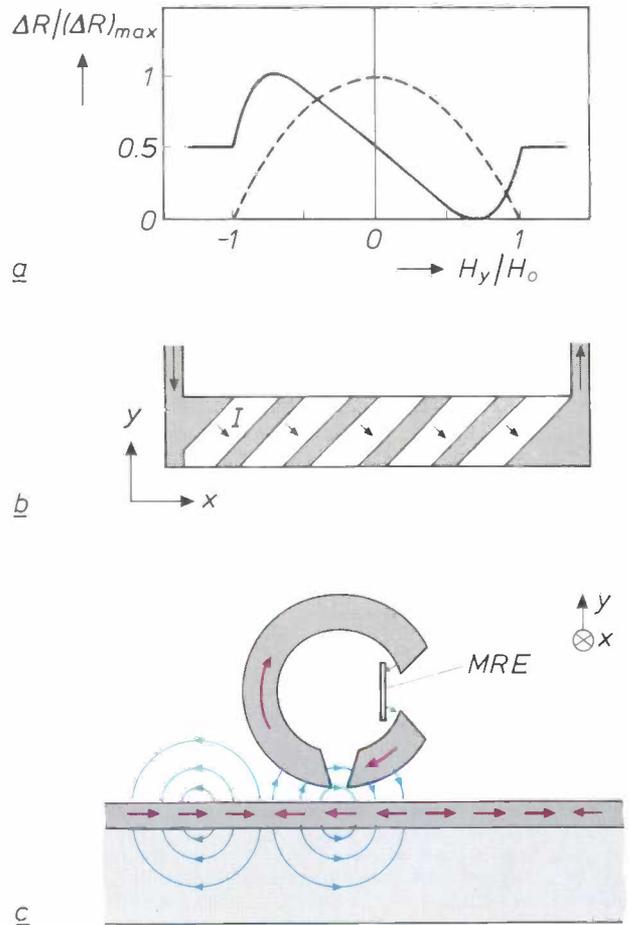


Fig. 4. *a*) The relative change $\Delta R/(\Delta R)_{max}$ in the resistance of a ferromagnetic strip as a function of the ratio of the magnetic field-strength in the *y*-direction to a reference field-strength H_0 [2]. The dashed curve refers to an untreated strip. The continuous curve, which in a wide range approximates to a straight line, refers to the case where a 'barber pole' structure of conducting material is applied to the strip. *b*) A strip treated in this way. The grey areas correspond to conducting material. For $H_y = 0$ the angle between the current *I* and the magnetization is 45° in the areas in between. *c*) Principle of a playback head with a magneto-resistance element *MRE* in the form of a 'barber pole'. A change in the magnetic flux in the yoke causes a change in the direction of the magnetization in *MRE*, and hence in the electrical resistance. The *x*-direction is perpendicular to the plane of the drawing.

strip does not just contain a single domain in which the magnetization has the same direction everywhere, but has different domains each with a different direction of magnetization. These are often known as Weiss domains and their common boundaries are called Bloch walls^[4]. In a rectangular strip the Bloch walls generally start from the corners; see *fig. 5a*. When a magnetic field is applied at right angles to the longitudinal direction, the various domains change in size, so that the Bloch walls have to change their positions. Owing to slight inhomogeneities in the material, there are discontinuities in the movements of the walls —

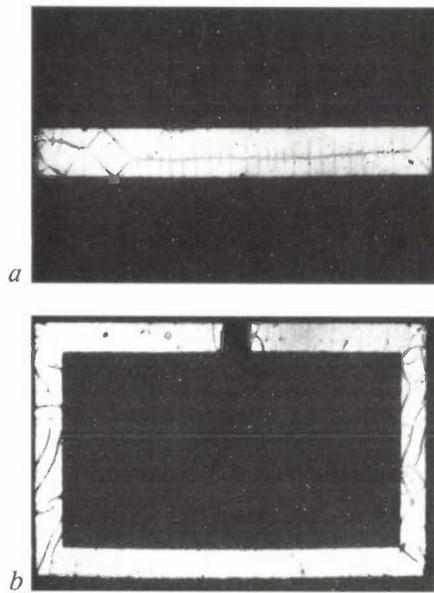


Fig. 5. The magnetic (or Weiss) domains *a*) in a rectangular magnetoresistance element and *b*) in a magnetoresistance element in the form of an interrupted loop.

the Barkhausen effect. As a result, 'Barkhausen noise' is superimposed on the output signal of a playback head.

We solved the problem of the Barkhausen noise by getting the Bloch walls to form outside the active part of the magnetoresistance element. Fig. 5*b* shows that this was done by making the permalloy strip in the shape of an interrupted loop. (The break is necessary to prevent circulating currents.) The corners where the Bloch walls originate thus lie outside the region of the 'barber pole'. Partly because a magnetic field is applied in the appropriate direction (horizontal in the figure) when the layer for the magnetoresistance element is deposited, there is only one domain at the location of the 'barber pole'. In the vertical parts there are more domains, but these do not cause any perceptible Barkhausen noise.

The r.f. bias field in the recording heads

In the magnetic recording of an analog speech signal an r.f. bias signal is superimposed whose frequency is much higher than the maximum frequency occurring in the analog signal. A bias frequency between 60 and 100 kHz is usually used. The bias signal induces an r.f. bias field in the recording head, and the bias ensures that the hysteresis of the tape material does not cause distortion of the recorded signal.

The bias current required for each recording head is about 100 mA. The total bias current for a 32-track CLS recording head is therefore about 3 A. A current

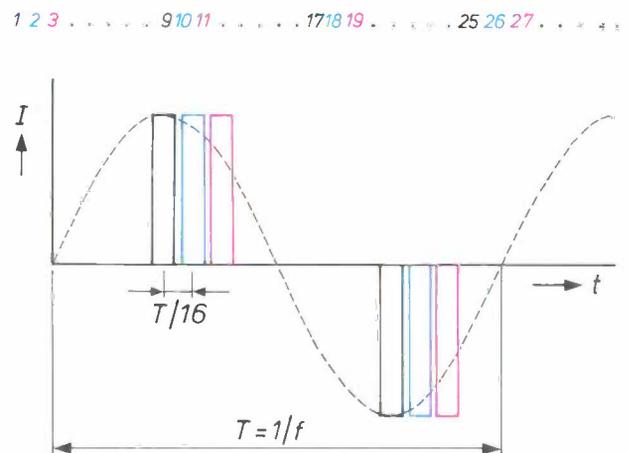


Fig. 6. The pulsed bias current supplied to the individual recording heads, numbered 1 to 32. *I* current. *t* time. The pulses shown in black are associated with the four heads with the black numbers, and similarly for blue and red, and so on. During any half-period of the original sinusoidal bias current (dashed line) there are eight successive pulses of the same polarity; in the next half-period there are eight more of the opposite polarity. *T* period. *f* frequency.

as high as this would raise the temperature of the recording heads above the Curie point of ferrite, 130 °C.

This problem was solved by using a pulsed current for the bias instead of the usual sinusoidal current and by not applying it to all 32 recording heads simultaneously. The pulse duration is only 10% of half a period of the original sinusoidal signal. Trains of positive pulses alternate with trains of negative pulses in successive half-period intervals; see *fig. 6*.

The analog speech signal is only applied to a recording head while a positive or negative pulse is present. The individual recording heads are divided into groups for pulse supply. If the heads are numbered from 1 to 32, a current pulse is applied first to heads 1,

[4] H. J. de Wit and K. Jager, Magnetic domains in amorphous alloys for tape-recorder heads, Philips Tech. Rev. 44, 101-109, 1988.

9, 17 and 25 (shown black). The heads 2, 10, 18 and 26 (blue) receive current pulses next, but with a delay of $1/16$ of a period, and so on. Then all the heads receive a succession of negative pulses, and so on. This procedure greatly reduces the amount of heat generated, and the heat is distributed more evenly over the surface.

so that each combination of recording head and corresponding counter-head can act as a kind of transformer. In all the good recording heads an r.f. current in the common conductor of the counter-heads should induce an alternating voltage at the same frequency and of amplitude above a certain threshold value. When a 32-track recording head is sawn from

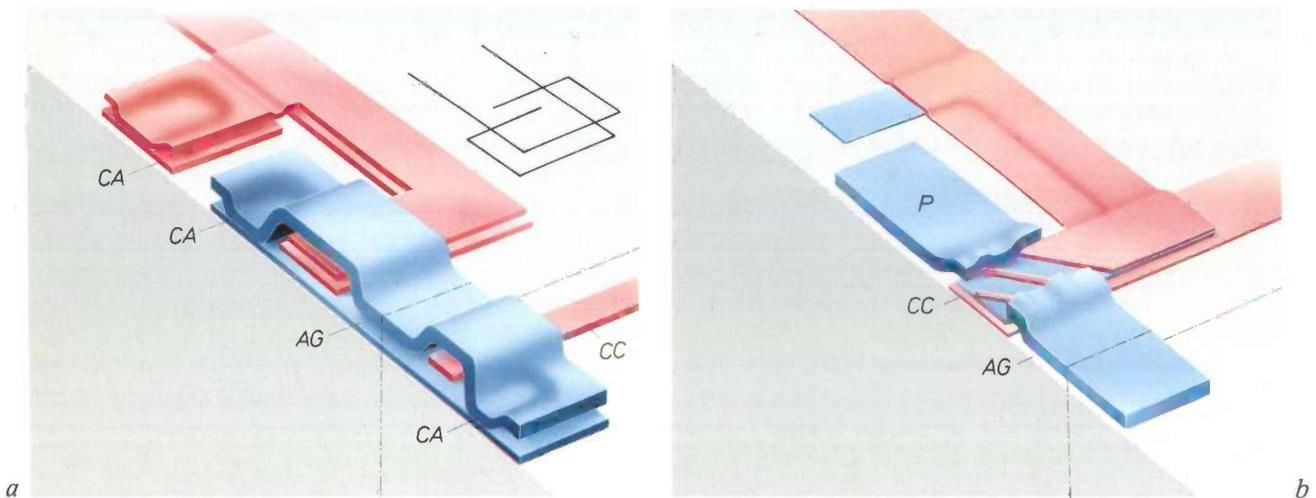


Fig. 7. *a*) A recording head and *b*) a playback head in thin-film technology. The heads have been cut through perpendicular to the surface of the magnetic tape. The chain-dotted line is the line of intersection with the plane of the tape, where the heads are sawn from the wafer. The part in (*a*) on the right below the chain-dotted line is the counter-head. Blue: permalloy; red: gold; grey: ferrite. The silicon dioxide has been omitted for clarity; it is assumed to be 'transparent'. *AG* head gap. *CC* common test conductor. *CA* contact areas of the two turns of the recording head and the flux guides, passing through SiO_2 layers between them. The two turns are shown schematically at the upper right in (*a*). The playback head only has an upper flux guide; the ferrite of the substrate acts as the lower flux guide. In this case the upper flux guide is separated by an SiO_2 layer from the substrate, since the relatively large area at *P* presents a low resistance to the magnetic flux.

The manufacture of thin-film magnetic heads

Thin-film technology

The yokes or flux guides of the heads shown schematically in figs 2*a* and 4*c* can be produced in thin-film technology by depositing thin films of permalloy on a ferrite substrate. The electrical conductors are formed by depositing thin films of gold, and the insulation between permalloy and gold is formed by depositing films of silicon dioxide. The head gap of $0.7\ \mu\text{m}$ therefore corresponds to an SiO_2 film of this thickness between the two flux guides.

A recording head built up in this way is illustrated schematically in fig. 7*a* and a playback head is shown in fig. 7*b*. The two turns of the recording-head coil lie one above the other, separated by SiO_2 . They make contact with each other through a hole etched in the SiO_2 . Each recording head has a 'counter-head', used for testing the heads while they are still on the wafer. All 32 counter-heads have a common conductor, *CC*,

the wafer, the counter-heads are destroyed by the saw-cut.

A 32-track playback head also has a common test conductor (*CC*) that runs underneath all the magnetoresistance elements. If this conductor is energized with an alternating current it produces an alternating flux in the flux guides. In a good recording head the resistance of the magnetoresistance elements should then also vary at this frequency. The test conductor can still be used after assembly of a complete 32-track recording head.

The films are deposited on the wafer by sputtering or by an electrochemical process. After application of photoresist, exposure through a mask and removal of the exposed resist, the film is etched away locally as shown in fig. 8. As we have said, the techniques are essentially no different from those used in IC technology^[5]. In general the manufacture of thin-film heads presents no difficulties in this respect and there is no

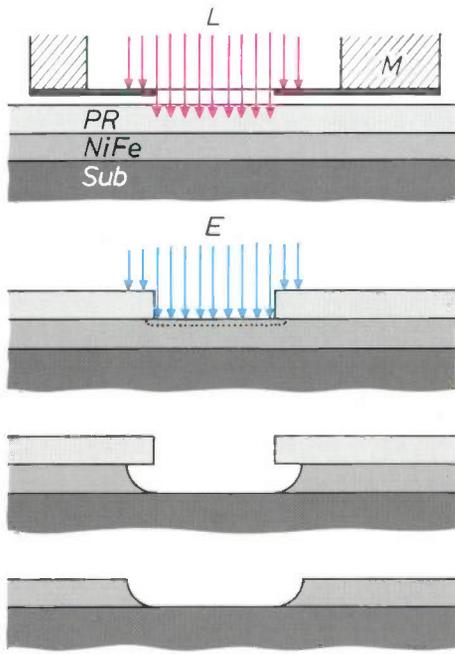


Fig. 8. Local removal of permalloy by means of a 'positive' photoresist, exposure and etching. *L* light. *E* etchant. *M* mask. *PR* photoresist. *Sub* substrate: ferrite or silicon dioxide applied in a previous process step.

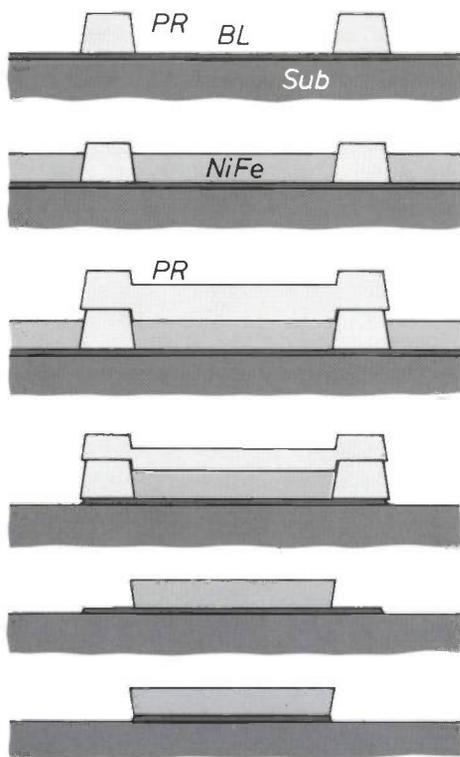


Fig. 9. The box-and-lid process. *BL* base layer. See also the caption of fig. 8. The substrate *Sub* is silicon dioxide from a previous step.

need for a detailed description here. We shall however look at a few techniques that are peculiar to the manufacture of thin-film heads: formation of a flux guide on an uneven surface, making patterns in gold and fabricating the magnetoresistance element.

In the recording heads the magnetic flux is greater than in the playback heads. This is why the recording heads have a lower flux guide of permalloy, which has a higher permeability and saturation flux than the ferrite of the substrate. In the playback heads, on the other hand, the ferrite acts as the lower flux guide. In the recording heads the lower flux guide is sputtered on to the flat base of the substrate. The upper flux guides in both playback and recording heads are applied on top of the SiO₂ and gold films applied prev-

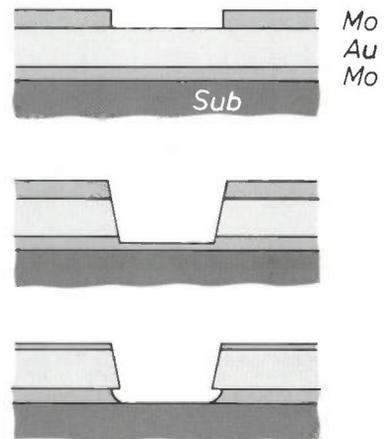


Fig. 10. Removal of gold, using a molybdenum mask, to leave a 'barber pole' structure. *Sub* substrate, in this case silicon dioxide.

iously, and therefore on a relatively uneven surface. The upper flux guide is relatively thick, several microns. Also, a layered structure is produced in the permalloy by periodically varying the ratio of nickel to iron. This 'laminated' structure is used to prevent eddy currents. Because of these various complications the upper flux guides have to be applied in a special electrochemical technique, the 'box and lid' process. This technique gives a much better side wall than an etching process, and the side wall is less affected by corrosion.

Fig. 9 shows the principle of the box-and-lid process. On a thin conducting permalloy base layer *BL* with titanium oxide beneath it for adhesion, walls of photoresist *PR* are built up, forming the side walls of the 'box'. Permalloy with periodically varying composition is then grown on the base layer. At the places where the flux guide is to remain, another layer of

[5] S. M. Sze (ed.), VLSI technology, McGraw-Hill, New York 1983.

photoresist is deposited (this is the 'lid'). Then the surplus permalloy is etched away, the photoresist is dissolved and finally what is left of the base layer is removed by a brief ion-etching process [6].

The gold for the electrical conductors is deposited by sputtering. Gold does not adhere well to silicon dioxide, however, and so an adhesion layer of molybdenum is applied first, again by sputtering; see *fig. 10*. A second layer of Mo is deposited on the gold. At the places where gold is to be removed an opening is made in the second Mo layer by wet-chemical etching through a mask of photoresist. The second Mo layer then acts as a mask for etching away the gold with a plasma of O₂ and Ar [7].

To protect the permalloy of the magnetoresistance element when making a 'barber pole' the etching process is stopped when about half of the lower Mo layer has been removed. The remainder is then etched away by a CF₄ plasma. The other electrical conductors are made by etching with O₂ and Ar alone, and the process continues until the lower molybdenum has been completely removed.

As mentioned earlier, the magnetoresistance element in a playback head takes the form of an inter-

rupted loop; see *fig. 5b*. The element need be only a few hundredths of a micron thick. The original layer, which will be etched later, can therefore be applied by sputtering. During the sputtering process a highly uni-

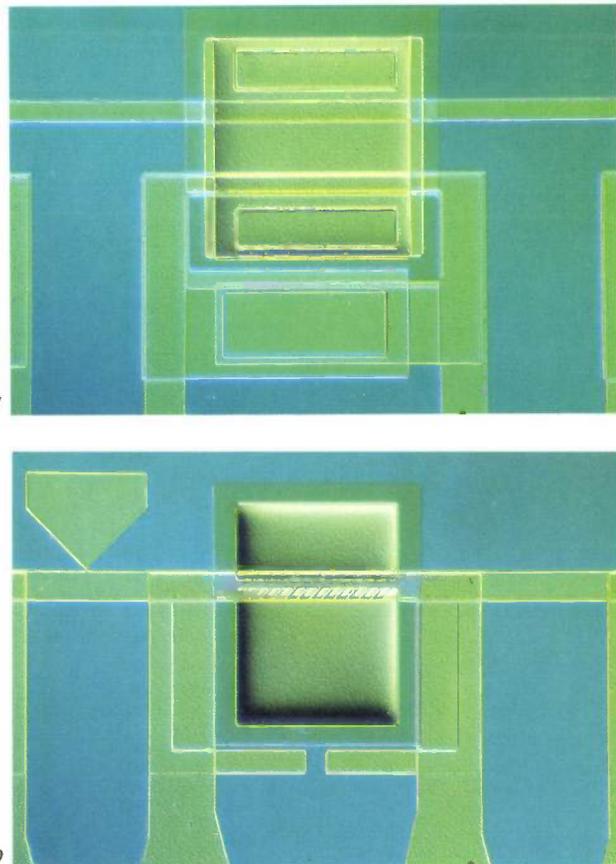
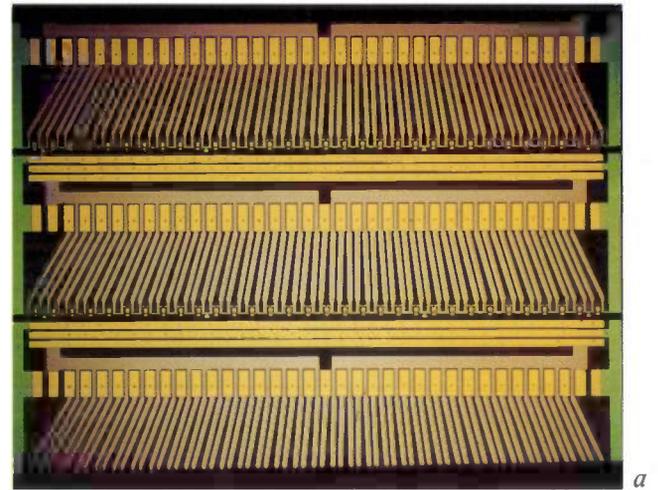


Fig. 11. Photomicrograph of *a*) a single recording head and *b*) a single playback head.

	B	A	
1	R _{geom} = 25.3 A _{geom} = -32.9 aantal foute kopjes = 0	R _{geom} = 25.1 A _{geom} = -33.7 aantal foute kopjes = 2	
2	R _{geom} = 25.6 A _{geom} = -32.5 aantal foute kopjes = 0	R _{geom} = 24.6 A _{geom} = -33.5 aantal foute kopjes = 2	
3	R _{geom} = .0 A _{geom} = .0 aantal foute kopjes = 32	R _{geom} = .0 A _{geom} = .0 aantal foute kopjes = 32	
4	R _{geom} = 26.0 A _{geom} = -32.3 aantal foute kopjes = 0	R _{geom} = 25.9 A _{geom} = -32.9 aantal foute kopjes = 0	
5	R _{geom} = 25.6 A _{geom} = -32.4 aantal foute kopjes = 0	R _{geom} = 25.9 A _{geom} = -32.7 aantal foute kopjes = 0	
6	R _{geom} = 25.3 A _{geom} = -33.0 aantal foute kopjes = 0	R _{geom} = 25.7 A _{geom} = -33.0 aantal foute kopjes = 0	
7	R _{geom} = 25.3 A _{geom} = -32.7 aantal foute kopjes = 0	R _{geom} = 25.2 A _{geom} = -32.9 aantal foute kopjes = 0	
8	R _{geom} = 24.8 A _{geom} = -32.7 aantal foute kopjes = 0	R _{geom} = 24.6 A _{geom} = -33.0 aantal foute kopjes = 0	
9	R _{geom} = 24.3 A _{geom} = -33.0 aantal foute kopjes = 0	R _{geom} = 24.3 A _{geom} = -33.2 aantal foute kopjes = 0	
10	R _{geom} = .0 A _{geom} = .0 aantal foute kopjes = 32	R _{geom} = 23.6 A _{geom} = -33.8 aantal foute kopjes = 3	

Fig. 12. *a*) Photomicrograph of three 32-track read heads still on the ferrite wafer. The indentations on the contact faces were made by test pins. *b*) Summary sheet produced by the computer from the automatic tester. Each frame contains the results of measurements on a 32-track playback head. The colour coding for the heads is: blue stripe — pass; dashed blue stripe — second harmonic of the signal transferred by the test conductor is too high; black stripe — interrupted magnetoresistance element; dashed black stripe — short-circuit in the element; red stripe — resistance of the element is outside the limits; green stripe — the amplitude of the transferred signal is too small.

form magnetic field is applied along the long axis of the magnetoresistance element to be formed in the etching step. In this way an 'easy' axis of magnetization is produced in the permalloy film.

The permalloy film is applied by r.f. sputtering, with high-frequency electric fields. Since the r.f. source and the load are not usually perfectly matched, reactive currents at the same frequency occur in metal near the plasma. We had to develop specially shaped permanent magnets to obtain a uniform magnetic field with the required direction at the ferrite wafers.

After the permalloy has been etched to produce the structure shown in fig. 5b, the next step is to deposit the 'barber pole' of gold and molybdenum on the element by the process described above. If the temperatures are too high, or if there are magnetic fields in undesired directions, the preferred direction of magnetization in the magnetoresistance elements may be lost. In this and later process steps the wafers must therefore be handled with extreme care.

After all the process steps have been successfully completed, a recording head and a playback head will look as shown in fig. 11.

Testing

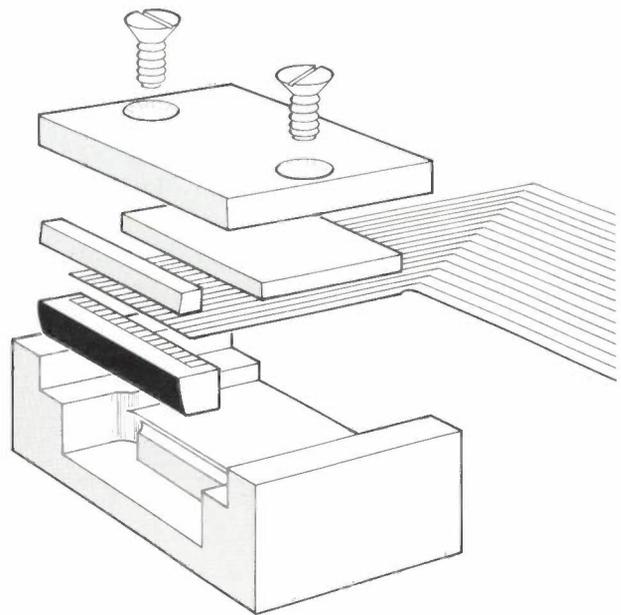
Fig. 12a shows part of a ferrite wafer with three 32-track playback heads. Each individual head is tested on the wafer before the multiple head is sawn off and assembled to form a head ready to be mounted in a system. The indentations made by the test pins in an automatic test instrument can be seen in the figure. A row of four test points (vertical in the figure) consists of, from top to bottom: two common connections to the test conductor, a common connection to the magnetoresistance element and an individual connection to this element.

The resistance of the magnetoresistance element and the magnetic transfer of a 2000-Hz test signal are measured for each head. A computer print-out from the automatic test unit lists the following figures for each 32-track playback head:

- the mean and the standard deviation for 32 results of the resistance measurement.
- the mean and the standard deviation for 32 results of the magnetic transfer measurement, and
- the number of heads that have passed the test.

The last figure is obtained by using specified pass values for the resistance and the magnetic transfer.

Fig. 12b shows a summary sheet produced by the computer for each measured ferrite wafer. A blue vertical line represents a head that has passed. Fourteen of the twenty multiple heads have passed. In two of the heads all the individual magnetoresistance elements are short-circuited. A summary of the pro-



a



b

Fig. 13. Assembly of a 32-track playback head. a) The part sawn from the ferrite wafer is clamped into an aluminium block with a ceramic 'tile' and a ribbon cable. b) The complete assembly of two 32-track recording heads and two 32-track playback heads, ready for mounting in a CLS logger.

[6] H. Dimigen and H. Lühje, An investigation of ion etching, Philips Tech. Rev. 35, 199-208, 1975.

[7] H. Kalter and E. P. G. T. van de Ven, Plasma etching in IC technology, Philips Tech. Rev. 38, 200-210, 1978/79.

duction results for each wafer can thus be obtained quickly. These results refer to laboratory-scale production, of course.

In the production of the 32-track recording heads the automatic tester is only used for measuring the resistance of the coils in the individual heads. Sample tests are taken of the magnetic-transfer of randomly selected individual heads by applying a 500-kHz signal to the test conductor in the counter-heads.

Assembly

The multiple heads that have passed the test are sawn from the ferrite with a fine diamond saw. The wafer is 2.4 mm thick. A 'tile' of ceramic barium titanate, 1.2 mm thick, is then bonded to the ferrite; see *fig. 13a*. This assembly is then polished to a rounded shape, with the required height of the head gap — the gap dimension perpendicular to the surface of the tape. The small triangles in *fig. 11b* and *12a* serve as reference marks for measuring gap height. Finally, the surface is lapped with diamond powder of decreasing grain size until the surface that comes into

contact with the tape is exceptionally smooth. This means that tape wear is extremely low.

Fig. 13 also shows how the ferrite-and-ceramic assembly with a flexible ribbon cable are clamped in an aluminium housing. The contact faces of the ferrite are pressed against the contact faces of the ribbon cable. After a final test the complete head is ready to be mounted in a Philips CLS logger.

Summary. CLS loggers (CLS stands for Communication-Logging System) used for logging spoken messages have to record or play back a large number of signals simultaneously at a very low tape speed. Multiple magnetic heads have to be used. Thin-film technology is used for fabricating 32-track recording heads that are different from the 32-track playback heads. The recording heads have the conventional coils, the playback heads have a magnetoresistance element. In the magnetoresistance element the resistance of a permalloy strip changes when the field in the magnetic 'yoke' of the playback head changes. The 32-track recording and playback heads are fabricated on a ferrite substrate by process steps like those used in IC technology: sputtering, electrochemical growth, photoresist deposition, exposure, plasma etching, ion etching, and photoresist removal. Test conductors in the heads enable tests to be carried out on the wafer. After all the tests have been carried out the multiple heads that have passed are sawn from the wafer and assembled to form a unit ready to be mounted in a logger.

1938

THEN AND NOW

1988

Car radios

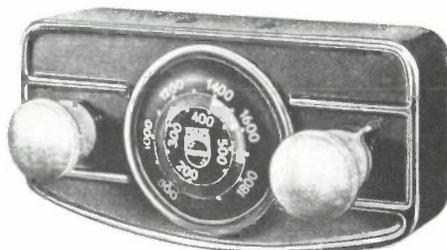


The first car radios, designed for long-wave and medium-wave reception, were built in the thirties. The valves and other components then available were far from small and the early sets, which sometimes had built-in loudspeakers, were so large (lower photograph^[*]) that they had to be fitted outside the reach of the driver. Remote control was achieved by using Bowden cables and a control unit (photograph right^[*]) fitted in the dashboard or underneath it.

In the past fifty years, there have been many radical changes in car radios. Transistors have been followed by integrated circuits. Car radios have become much smaller, they are easier to operate, and the performance is much better. There are so many new facilities — FM, FM-stereo and cassette players — that we speak of 'in-car entertainment'.

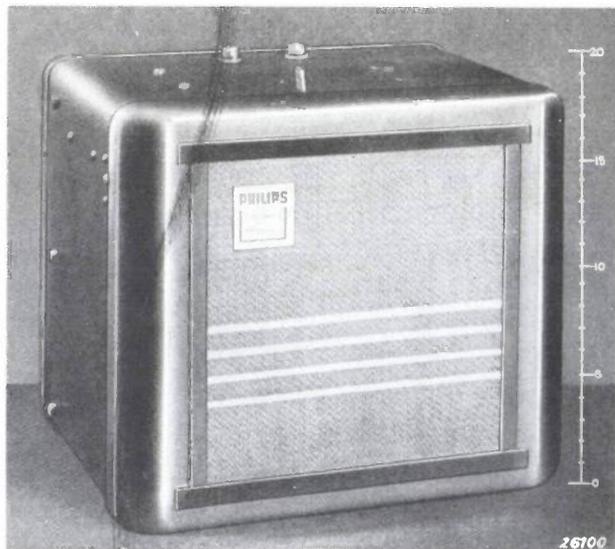
The most advanced car radio/cassette combination today is the Philips DC 682 (colour photograph). It

makes use of coded digital information added to radio transmissions in the 'radio data system' (RDS) developed by European companies and broadcasting authorities. Philips Research have played a leading part in this development and Philips were the first to



put an RDS car radio/cassette combination on the German market. RDS will soon be introduced in other countries as well.

The DC 682 incorporates the primary RDS features: it displays the station selected, switches automatically to the strongest transmitter with the desired programme, and traffic information will override the broadcast signal. It also has many features from other recent Philips car radios which were unthinkable in 1938: preselection from eighteen stations, digital tuning, autostore for automatic tuning to the six strongest FM and MW stations, 2×20 W or 4×7 W output with fader, line-out, anti-theft security code and retractable unit, and autoreverse cassette deck with Dolby noise reduction, metal/chrome tape selection and Music Sensor System for track repeat and skip. RDS will include many more features in the future, like paging, clock time, radio text and programme type (e.g. sport, jazz).



[*] From Philips Technical Review, April 1938.

Applied research — the source of innovation in consumer electronics

S. van Houten

On 1st June 1988 Dr S. van Houten, a member of the Board of Management of N.V. Philips Gloeilampenfabrieken, gave an invited speech in Nanjing in the People's Republic of China. In his speech Dr van Houten gave a short introduction to the Philips company, then used examples to show how important the role of scientific research is in bringing out a wide variety of electronic consumer products.

China may be distant, but its people are rapidly learning to master the latest technologies, and the speaker was privileged to address a large and attentive audience.

The article below is the text of his speech; we have merely added references and figure captions.

A picture of Philips

The company to which I belong is called Philips Electronic Industries. Our headquarters are located in Eindhoven in the Netherlands, a provincial town situated roughly midway between Amsterdam and Brussels — if that is of any help to you.

Philips Electronic Industries is a major electronics company — in fact number six in the world in terms of sales (*fig. 1*). In 1987 sales totalled 28 billion dollars at current rates. Philips employs over 330 000 people in 60 countries throughout the world. About two-thirds of our activities are concentrated in Western Europe, one quarter in North America. The remainder is divided more or less equally between Latin America and the countries of Eastern Asia.

Among the latter, our activities in the People's Republic of China are growing fast; they are generally carried out by joint ventures in which government agencies participate. Under this scheme we jointly manufacture car radios and radio recorders which are to be sold under the Philips brand name throughout China. Agreement has been reached on the joint manufacture of fluorescent lamps and colour picture tubes. Innovative products such as Compact Discs,

CD players and optical cables will in the future be manufactured under joint ownership. Video cassette recorders and integrated circuits will follow.

Philips was started just under a century ago, in 1891, by two brothers named Philips as a light bulb factory. I show a historic picture here: the workers of that first year (*fig. 2*). About 40 years later, the company diversified into gas-discharge lamps, X-ray equipment, radio, gramophones and measuring instruments. Electric shavers and household appliances

			Sales (US\$ × 10 ⁹)
1	IBM (International Business Machines)	U.S.A	51.3
2	General Electric	U.S.A	35.2
3	AT&T (American Telephone and Telegraph)	U.S.A	34.1
4	Matsushita	Japan	26.5
5	Hitachi	Japan	22.7
6	Philips	The Netherlands	22.5
7	Siemens	W. Germany	20.3
8	Samsung	Korea	16.5
9	Toshiba	Japan	15.0
10	CGE (Compagnie Générale d'Électricité)	France	11.7

Dr S. van Houten is a member of the Board of Management and the Group Management Committee of N.V. Philips' Gloeilampenfabrieken.

This speech is published by permission of the Corporate External Relations Department, Philips International B.V.

Fig. 1. The ten largest producers of electronic equipment in 1986, in order of sales (source: Fortune).



Fig. 2. The Philips personnel in the year the company was founded (1891). Today, nearly a century later, there are more than 330 000 Philips people in some 60 countries.

were to follow, along with telephone exchanges. Television supported an unprecedented growth in the post-war era. Optical discs carrying sound, picture and data are among our latest inventions.

This means that Philips now covers a wide range of disciplines. For all new fields of activity, the essential components are manufactured in our own plants. We express this by saying that our company is vertically integrated.

Research at Philips

The successful diversification of our company could not have been achieved without a firm base of fundamental knowledge and technical know-how. Our corporate research has played an essential role in the history of our company. Next year it will be 75 years^[1] since the first scientist was recruited; he was told to conduct research into the gas-discharge phenomenon. Today our Research Laboratories in the Netherlands employ 2500 people; roughly the same number is employed in seven other laboratories scattered all over the world. All of them are doing fundamental research: product development is done by another 20 000 people. Philips spends a total of more than 8% of its sales on Research and Development — roughly \$2.5 billion each year.

As is apparent from the wide variety of Philips products, our research covers a great number of disciplines. Thus we are fortunate in being capable of a multidisciplinary approach to modern systems as modern technology depends increasingly on the combined use of several disciplines. Materials science, fundamental physics, electrotechnology, chemistry and software are combined in products which are to be found everywhere, in hospitals, workshops, banks and — the home.

Modern fluorescent lighting

Fluorescent lamps^[2] are a good example. They are an offspring of the gas-discharge lamps which our first research scientist was told to investigate almost seventy-five years ago. The gas discharge generates ultraviolet radiation which causes specially selected phosphors to emit visible light. The first such lamps, which were tubular, came on to the market in the 1930s. From then on, a continuous process of optimization took place. The mercury-vapour discharge was studied in detail, more efficient phosphors were found, the emission spectrum was adjusted to improve colour rendering, the ignition circuit (the 'ballast') was miniaturized by the introduction of electronics. Finally, our PL* and SL* economy lamps were introduced: these are roughly the same size as incandescent lamps but are four times more efficient (fig. 3).



Fig. 3. In recent years developments in fluorescent lamps have led to compact lamps with more than four times the efficacy and five times the life of comparable incandescent lamps. Examples are the PL* lamp (a) and the SL* lamp (b). The SL* lamp has a built-in ballast and the same base as conventional incandescent lamps, so that existing fittings can be used.

[1] A special multiple issue of Philips Technical Review on this 75th Anniversary will appear in 1989.

[2] A. G. Jack and Q. H. F. Vrethen, Progress in fluorescent lamps, Philips Tech. Rev. 42, 342-351, 1986.

TV technology

TV technology always has been an important area of research with Philips^[3]. Let us have a look at the video screen; it occupies a key position in consumer electronics. The existing technology, based on electron beams, is still advancing. But new developments are coming on to the market, namely liquid-crystal screens.

The colour picture tube

The electron optics of the conventional colour picture tube have reached a high level of precision, the luminescence of the phosphors has increased over the last few years and new manufacturing techniques have been developed for the electron gun. An important development of the last few years is the 'flat square' tube which has a flatter, more right-angled screen (fig. 4). The tube is made of thicker glass; special strength calculations were needed to ensure a safe design. Thanks to perfected electron optics the colour registration right into the corners of the picture is outstanding.

Projection television allows screen sizes of over 1 metre, diagonally (fig. 5). These larger sizes give an



Fig. 4. Modern television picture tube, type 45AX. This is a 'flat square' tube, with the screen as flat and rectangular as possible and the total tube depth as small as possible. The tube shown here also has an extremely thin neck (29.2 mm), so that the deflection of the beam requires less energy.



Fig. 5. To obtain a larger picture than with a 'direct-view' picture tube of reasonable dimensions and weight, a projection television set can be used, as shown here. This contains three projection tubes that each produce a subpicture in one of the basic colours red, green or blue. The three subpictures are projected on to the screen exactly in register.

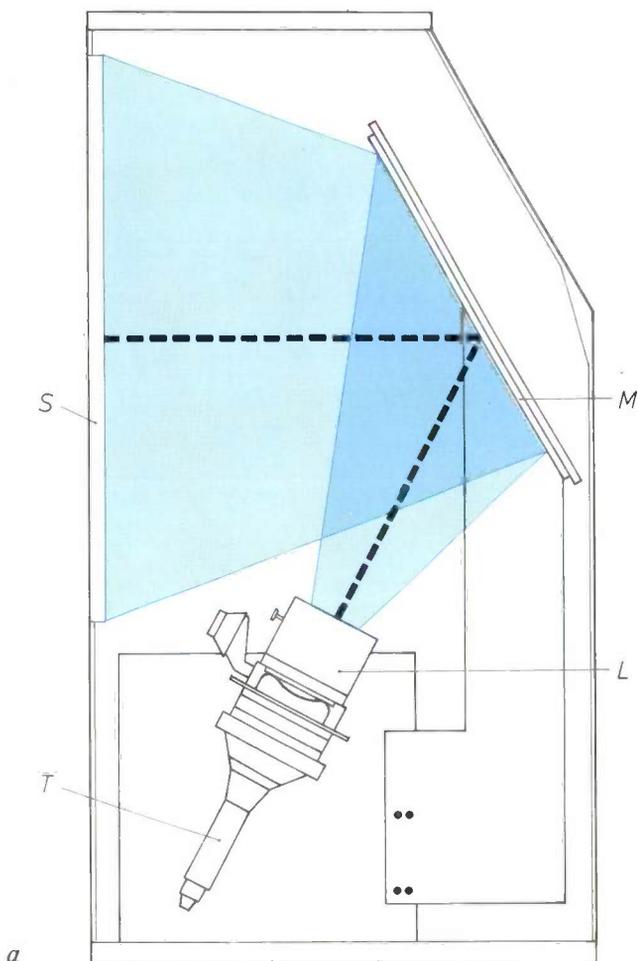
enhanced viewing experience. Three projection tubes provide the colour picture (fig. 6a). Projection cathode-ray tubes have been developed with electron guns which emit an extremely intense electron beam, and phosphors which are subject to 50 times the load of a normal picture tube. 100 watts are continuously applied to the phosphor screen; this necessitates liquid cooling. A special high-grade lens system with aspherical lenses has been developed for the projection optics; this is coupled to the screen via the coolant (fig. 6b).

The projection optics have to meet strict quality requirements. At the same time, however, they have been designed so that they can be produced at a reasonable cost. In projection TV we see once again the many different disciplines involved and the conflicting requirements that are to be met. Once again it is the task of the research department to find the solution^[4].

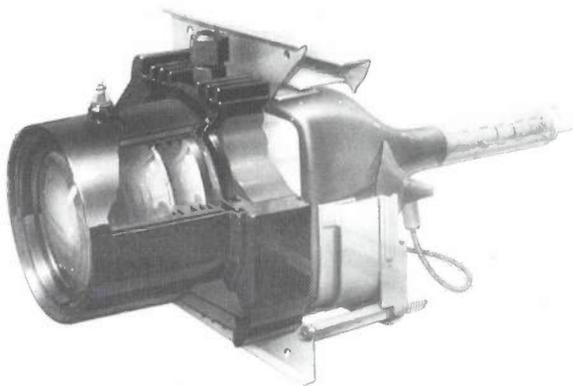
LCD displays for direct-view and projection television

Research also finds completely new ways. Liquid-crystal displays or LCDs will probably be familiar to you from your wrist watch. Their application to colour TV is a recent achievement, which has the advantage that it does away with the bulky and power-consuming picture tube. A pocket-size TV set as shown in fig. 7 becomes feasible. Larger displays can be hung on the wall like a picture.

Liquid crystals are peculiar liquids. They consist of elongated cigar-shaped molecules. The plane of polarization of the light propagating through the liquid follows the arrangement of the molecules.



a



b

Fig. 6. a) Light path (in blue) in one of the three projection tubes in a Philips projection television. The light is not directly incident on the back of the projection screen *S*, but is reflected by a mirror *M*. Each projection tube *T* is connected to a lens system *L* consisting of four lenses, some of them aspheric. The connection is made via a liquid that also cools the front of the tube. b) This is called a liquid-cooled, liquid-coupled tube/lens system.

When using liquid crystals for visual display purposes they are inserted between two narrowly spaced parallel glass plates which are covered on the inside with transparent electrodes. In the absence of an electric voltage between the electrodes the molecules assume a twisted configuration, as shown in *fig. 8a*. There is a polarizer on top of the assembly. The plane of polarization of the light falling through the glass plates is twisted in the same way as the molecules. We have designed the unit to give rotation through 90° . A crossed polarizer is present at the output side of the unit and, thanks to the 90° rotation, light will be transmitted. Application of an electric voltage between the electrodes causes the molecules to align axially (see *fig. 8b*); they will no longer influence the plane of polarization of the light and the crossed polarizers will cancel the light.

It is a long way from this principle to a TV display in which a great number of picture elements have to be switched independently. Multiply that number by three for the three different colours. In the Philips display, electrodes are replaced by vapour-deposited



Fig. 7. The use of liquid-crystal displays (LCDs) opens new possibilities for television. The colour television receiver shown here has a screen with a diagonal of 7.5 cm (3 in) and requires only a few watts of power.

[3] K. Teer, Looking back at distant vision: television technology from 1936 to 1986, Philips Tech. Rev. 42, 297-311, 1986.
 [4] Three articles on research into subjects of direct interest for the development of projection television will shortly appear in this journal. These articles deal with interference filters, phosphors and electron guns.

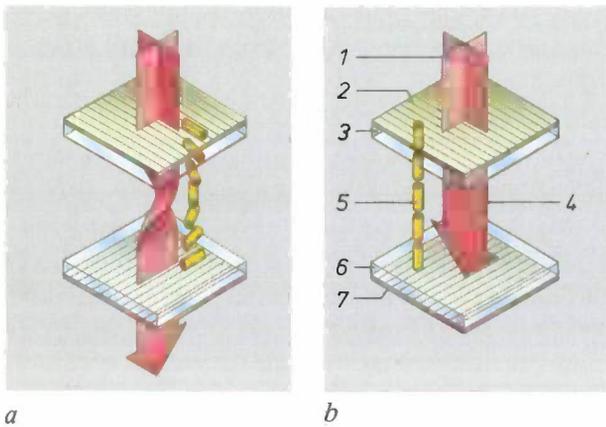


Fig. 8. Basic principle of operation of a liquid-crystal display (LCD). *a*) If no voltage is applied, some of the incident light is transmitted through the LCD. *b*) If a voltage is applied, no light is transmitted. 1 incident light; 2 polarizer; 3 glass; 4 polarized light; 5 liquid-crystal molecules; 6 glass; 7 polarizer.

thin-film transistors. Here too, close cooperation between different disciplines is essential: materials science, cell and circuit design, production under strictly controlled conditions.

Backlighting is required when viewing your LCD colour display. Large-area, energy-efficient light sources are needed; here our lighting expertise is of great value.

Projection TV using LCD light modulation is under investigation. The LCD is flat and thin; it could take the place of the slide in a slide projector.

HDTV

Currently, a debate is going on about television with greater picture resolution: High-Definition TV or HDTV as it is known. There is a call for pictures with more lines, for example 1250 instead of 625 — and thus with more information. Projection on larger screens is then possible, as is the introduction of a broader 'CinemaScope'-type picture. Transmission channels with a larger bandwidth are needed for the broadcasting of these pictures. A number are available thanks to the advent of broadcast satellites having a fixed position high above the earth's equator. The Federal Republic of Germany has launched one. Unfortunately it is not functioning properly: one of the two solar panels failed to unfold in space. It is now a question of waiting for a similar satellite to be launched by France.

For television broadcasts over the earth's surface and for recording on magnetic tape it is necessary to limit the bandwidth. There is ample room for this because, in normal TV signals, the same information is transmitted over and over again. You see, large parts of consecutive TV pictures are often the same, and

changes take place relatively slowly. For bandwidth reduction there are sophisticated digital signal processing techniques which use mathematical transformations. This is a typical research matter.

Of course, a standard is also necessary for HDTV broadcasting, so that the owner of the television set can receive broadcasts from any transmitter. We would really like to see a world standard. A complicating factor is that the picture frequency of broadcast television is linked to the mains frequency. In the United States, Canada and Japan that is 60 hertz, in most other areas, including Europe — and China — it is 50 hertz.

In Europe cooperation between Philips and other electronics manufacturers has produced a standard called MAC, which stands for Multiplexed Analog Components. One of its features is that the colour and brightness information are not transmitted simultaneously, but the colour first and then the brightness. Fig. 9 shows how this works; after decoding there is again a colour picture. In this way, interference, known as cross-colour and cross-luminance, is

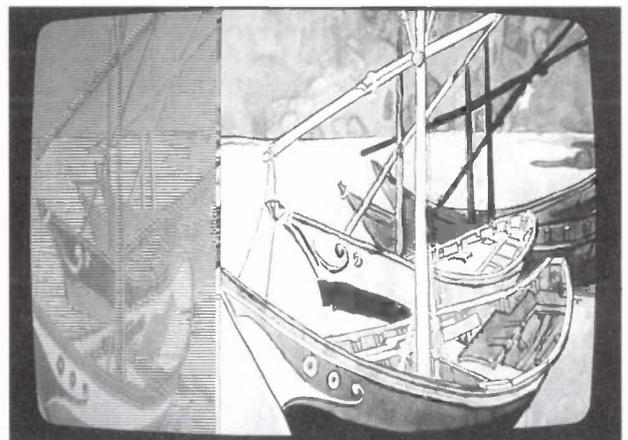


Fig. 9. Picture obtained when a MAC signal is displayed without special decoding, as if it was an ordinary television signal. In every line period a small part is first used for transmitting the chrominance information, and the rest is used for the luminance information. Since there are two chrominance signals (U and V), U -information and V -information alternate from line to line. A colour picture of very high quality can be obtained with a MAC decoder, because all the various undesirable interactions between chrominance and luminance information are intrinsically avoided with this system. (The line-flyback periods are used for sound transmission; this is not visible here of course.)

avoided. The MAC system is compatible with the current standards and can therefore be applied alongside the current system in a transitional period. This prevents all TV receivers becoming suddenly unusable upon the introduction of HDTV^[6].

Storage

Consumers are not satisfied by picking up radio and television signals from the air. There also is a great demand for recorded music and television. Storage devices and storage media are an important sector of the consumer market.

Magnetic tape recording

For many years the gramophone record and the magnetic tape have fulfilled the existing need. Magnetic recording on tapes has developed to a high degree of sophistication. Much of the gain in information density on the tape surface is attributable to the ongoing improvement in tape quality and the improved characteristics of the recording heads. This has enabled us to attain a satisfactory recording quality with ever-decreasing tape speeds.

A milestone in sound recording for the consumer public was the introduction in the 1960s of the Philips Compact Cassette, which made use of the low tape speeds that had just become acceptable. Recording video on tape at that time was still the preserve of the professional community. Once again it was Philips who first introduced the video cassette recorder for home use.

Throughout the years magnetic recording has remained one of the key topics of our research. Apart from improving the physical characteristics of the tape and tape heads^[6], and developing more accurate mechanics and servocontrol systems, recording density on the tape has been increased by proper signal processing. Most tape recording is done by writing analog signals. It is only recently that digital tape recording of sound has become commercially viable. Digital video recording is available in our laboratories.

In the meantime, optical recording media have been introduced. Compared with these, magnetic tape recording trails far behind in terms of recording density. The audio Compact Disc carries about one million bits per square millimetre, a digitally recorded audio tape four times less. Secondly, magnetic tape does not allow rapid access. However, it has the advantage of bulk storage — it can provide hours of playing time.

Optical recording media

In fact it was IC technology, in which intricate details of optical fineness are repeated flawlessly over and over again, which, in our laboratories, gave rise to the idea that the same means could be used to record information on discs. 1972 saw the introduction of the LaserVision video disc on which video information is recorded in analog form. The LaserVision disc can store up to 40 000 still pictures; in the well-publicized Domesday Book project it has been used to

record data about all of Great Britain for educational purposes. Meanwhile, the art of digital signal processing progressed at a rapid pace, and 1982 saw the birth of the Compact Disc — also an invention by Philips. It carries audio information in digital form and has set a new standard of high-quality sound recording.

The Compact Disc is more than just an electronic device. It is part of a system, the Digital Audio Compact Disc System^[7]. In consumer electronics too the trend is one of moving away from stand-alone products and more towards systems. The Digital Audio Compact Disc System first of all encompasses many technical disciplines: the laser, the chemical process of the disc manufacture, servomechanisms for focusing, tracking and motor control, sophisticated optical systems, digital signal processing including error correction, etc.

Secondly, a *standard* is also part of it. It must be possible to play a CD from one manufacturer on any other manufacturer's player. And given that today's market for consumer electronics is a *world* market, the standard must be a *world standard*. Throughout

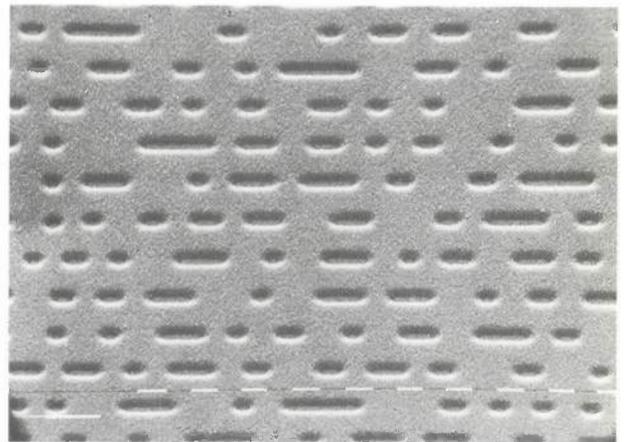


Fig. 10. The sound information on the Compact Disc is recorded as a succession of pits in a spiral pattern. The measurements here are microscopically small: the pits are only $0.6\mu\text{m}$ wide, $0.12\mu\text{m}$ deep and the length varies between 0.9 and $3.3\mu\text{m}$.

[6] M. J. J. C. Annegarn, J. P. Arragon, G. de Haan, J. H. C. van Heuven and R. N. Jackson, HD-MAC: a step forward in the evolution of television technology, *Philips Tech. Rev.* 43, 196-212, 1987.

[6] H. J. de Wit and K. Jager, Magnetic domains in amorphous alloys for tape-recorder heads, *Philips Tech. Rev.* 44, 101-109, 1988; J. P. M. Verbunt, Laboratory-scale manufacture of magnetic heads, *Philips Tech. Rev.* 44, 151-160, 1988; M. G. J. Heijman, J. H. W. Kuntzel and G. H. J. Somers, Multiple-track magnetic heads in thin-film technology, this issue, pp. 169-178.

[7] See our special issue on the Compact Disc: *Philips Tech. Rev.* 40, No. 6, 149-180, 1982.

the world the digital information must be recorded in the pattern of small pits (*fig. 10*) on the disc using the same code. We, when developing the Compact Disc, considered the issue of standardization as very important. To establish a world standard it is necessary to work together with others; we chose to cooperate with the well-known Japanese company Sony, which also specializes in consumer electronics. Sony made major contributions to the coding on the disc; the code has an extraordinary capacity to correct errors which arise when the disc is being read.

The capabilities of the Compact Disc system extend far beyond sound reproduction. The Compact Disc Video stores video, which is encoded in the Laser-Vision format alongside fully digital sound (*fig. 11*). The Compact Disc Read-Only Memory or CD-ROM is a fully digital storage medium for computer read-out; the Compact Disc Interactive or CD-I is derived

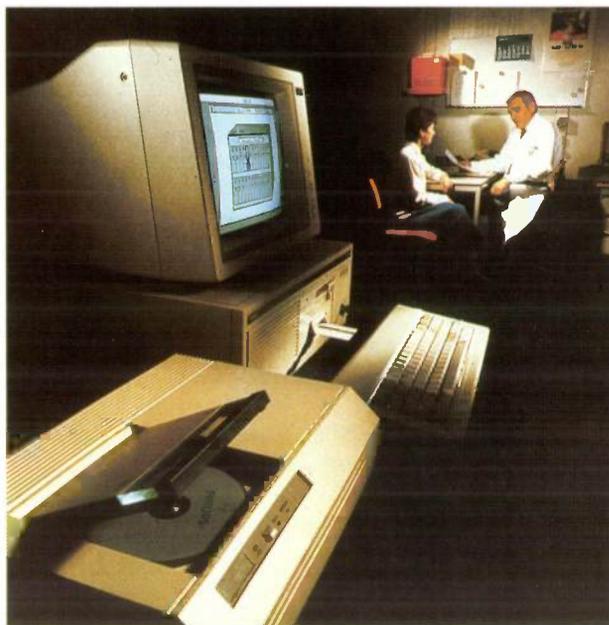


Fig. 12. Two highly promising versions of the Compact Disc are the CD-ROM and the CD-I, which each contain some 600 megabytes of information on a disc with a cross-section of about 12 cm. This photograph gives an impression of the use of a CD-ROM. For both CD-I and CD-ROM screens have an important part to play in displaying the information.



Fig. 11. The well-known 'silver' Compact Disc (second one down in the drawing) is slowly acquiring a tribe of audio and video relatives. From top to bottom: the CD Single, the CD, the CD Video Long-play and the LaserVision video disc. All of these discs can be played on a single universal combination player.

from CD-ROM for purposes of interactive communication with a personal computer (*fig. 12*); one existing CD-I contains an entire encyclopaedia.

So far, I have mentioned pre-recorded discs only. The user may wish to record his own data. Research shows different techniques to make this possible^[8]. All of them rely on local heating of the disc surface by a laser. In one technique, the laser melts small holes into a thin metallic layer applied on to the disc. The process is irreversible; we call this process 'Write Once, Read Many times'; it is very useful in many computer applications.

In a second technique, the laser produces amorphous spots in a crystalline layer; just as the holes

of the first example, the amorphous spots can be detected by scanning the disc with a laser beam. This process is reversible; by giving the disc an overall thermal treatment, crystallinity is restored. A third technique is based on a magnetized layer in which the direction of the magnetization is locally inverted by the heat of the laser beam. The state of magnetization is read out optically. This last technique also is reversible; it is the one which generally is referred to when one speaks about 'erasable Compact Disc'.

Obviously, exploring these different techniques requires very fundamental materials research. The life expectancy of the recording is an important issue.

Another aspect is the intricate digital processing to which the signals are subjected in order to exploit the recording media optimally. This is valid for magnetic as well as for optical recording. For both the challenge will be to record High-Definition Television, which generates a bit stream not of 140 megabits per second, as standard television does, but of 560 megabits per second. Intricate bandwidth-reduction techniques are necessary to meet this new challenge. They will be feasible only if Very Large-Scale Integrated circuits will be available to reduce the signal processing hardware to reasonable proportions.

[8] G. E. Thomas, Future trends in optical recording, Philips Tech. Rev. 44, 51-57, 1988.

ICs

And thus we arrive at what is perhaps the mainstream of research in today's electronic industries — integrated circuits. Hardly any technology has shown a faster development than that of circuit

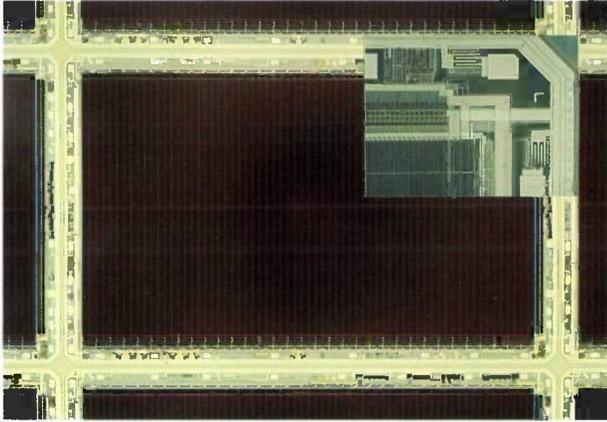


Fig. 13. In the present state of technology millions of transistors can be integrated on a single chip. The photograph shows a static random-access memory with a capacity of more than a million bits (the 1-megabit SRAM). The smallest details measure only $0.7\ \mu\text{m}$, while the entire chip has an area of about $94\ \text{mm}^2$; an enlarged view of a portion of the IC is given at the upper right-hand corner to show the details more clearly.

integration. It is only 20 years ago that Philips produced its first commercial integrated circuit; it was a three-transistor amplifier circuit for a hearing aid, realized on 0.6 square millimetres of silicon.

Today, for contrast, we are happy to have just produced in our laboratory first samples of an integrated circuit having about ten million transistors on an area of no less than 90 square millimetres. To accommodate so many transistors on the surface of one chip it is necessary to reduce the dimensions of the individual transistor. We aim at structural details as small as 0.7 micron, smaller than anything that is on the market today. The small dimensions have another advantage: small transistors are faster than large ones.

Our first goal is a one-megabit Static Random-Access Memory or SRAM incorporating one million six-transistor memory cells (fig. 13). Fabricating error-free ICs of this size in the sub-micron technology puts extreme requirements on processing conditions; as an example, it requires highest-class clean rooms. We have had to build special laboratories for this project. One of the buildings (fig. 14) is the pilot plant. As you will notice it looks like the well-known dual-in-line IC package. The clean rooms occupy



Fig. 14. An entirely new centre for studying and manufacturing ICs with details smaller than a thousandth of a millimetre — 'submicron details' — has appeared on the Philips Research Laboratories site in Eindhoven in recent years. The large building in the foreground is the pilot-production plant. The building behind it by the lake houses the research and design department.

about 15% of the floor area, the remainder being required for air conditioning and dust filtering. The building next to it houses hundreds of IC designers [9].

The costs of a research project like this are tremendous; in fact, they are expressed in billions of dollars rather than millions. These vast sums are beyond the means of an individual company; we are happy to cooperate with Siemens of West Germany. The project is of great significance for all of Western Europe, as it will give Western Europe a position in the forefront of modern IC technology. For this reason the project receives support from the West-German and Dutch governments.

Nor will this be the end of IC developments. In our laboratories, we have created transistors of even smaller dimensions, with 0.4 and 0.2-micron details. They have been configured by direct electron-beam writing on the silicon surface (*fig. 15*). They will be components of the future, even larger ICs storing tens of millions of bits or doing an equivalent amount of signal processing.

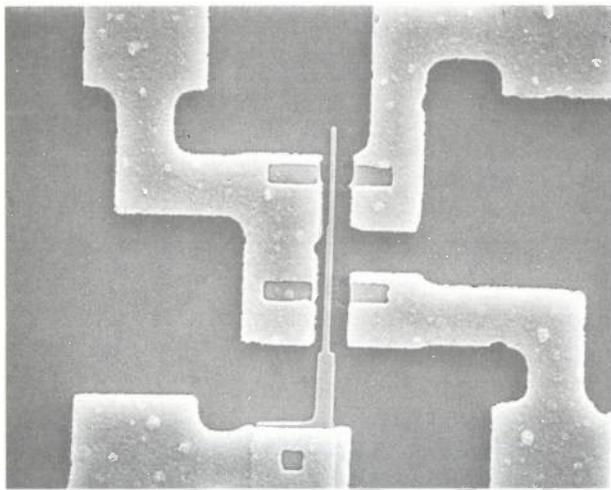


Fig. 15. The precursor of the next generation of ICs with detail dimensions from 0.2 to 0.4 μm . In this photograph two MOS transistors can be seen with common source (connection top left), common gate (connection lower left) and separate drains (connection top right and lower right). The thin 'needle' at the centre is the actual gate and has a width of 0.5 μm in one transistor and only 0.4 μm in the other. These small dimensions are achieved by using an electron-beam pattern generator. The relatively thick connections for the two transistors are of aluminium and are made by photolithography.

Philips is the largest European manufacturer of ICs in current technology. In bipolar ICs for analog signals it is even one of the biggest in the world. Analog ICs have been a speciality of Philips since many years; they are used in consumer products like TV sets and hi-fi audio, as well as in domestic appliances. In electric shavers they check the charge level of the batteries.

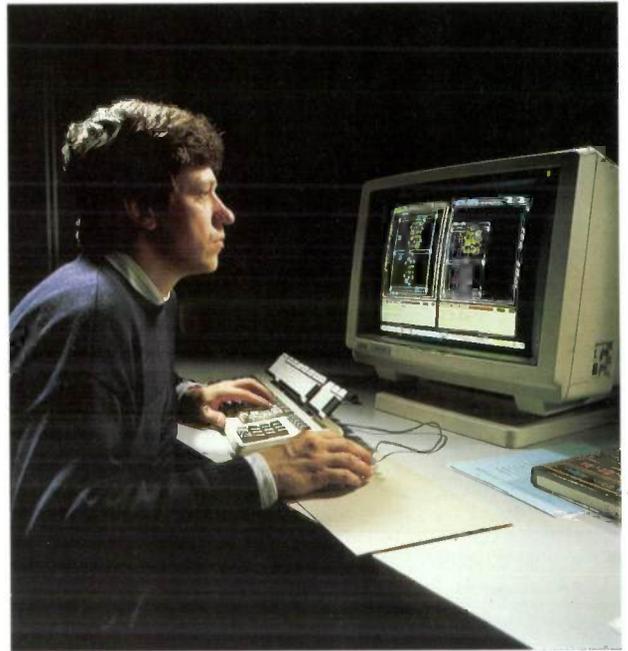


Fig. 16. The computer is becoming more and more vital to the automation of the design process for integrated circuits. There is no other reliable way of managing the enormous complexity of the latest semiconductor circuits in an acceptable time.

But let me return to the ever-larger digital ICs that are being developed for signal processing and for mass storage. As an application, I have mentioned the digital signal format of the CD. It would have been impossible only a few years before. I have mentioned high-definition TV and digital recording. They ask for many megabits of storage capacity, to store entire TV frames. This requires several square centimetres of silicon realized in the most advanced processes.

It is, however, not enough to have the newest technology available. One needs system concepts and designs to use it. It is a tremendous task to translate one's ideas into an electrical circuit on a chip. Standard design methods fail with the growing complexity. Too many designers are needed. It becomes impossible to survey a total design, and this will cause so many errors that the process will fail — or the design process will simply take too much time. In order to overcome these difficulties the total design process must be upgraded — a task comparable in size to the upgrading of the technology.

Of course, computers are extensively used in circuit design. Large-scale drawings are plotted of IC layout. But these methods no longer work. With the increasing scale of the circuits the size of the drawings will exceed the floor area of our buildings. What we need is a higher level of design automation. We are working towards silicon compilation — upon specification of the desired functions the computer produces the IC layout (*fig. 16*).

Concluding remarks

Let me conclude by three remarks.

- Modern consumer products are research-intensive. The time is gone that they were simple products and derived their technology from elsewhere, i.e. from professional fields.
- The distinction between the consumer and the professional field disappears or is even reversed. Modern silicon technology and design methods make it possible to integrate highly sophisticated functions into consumer goods. There are only two fields of industry which are of a sufficiently large economic scale to support the high costs of the new technologies — the computer industry and consumer electronics. Small-scale professional production will not be able to bear the high expense. As a result, we encounter a degree of sophistication in today's consumer goods which earlier was found only in professional products. As examples I mentioned the error-correcting code of Compact Disc or the bandwidth-reducing algorithms

on behalf of transmission and recording of high-definition television.

- These consumer goods are to be handled by non-specialists. In spite of their sophistication, they should be easy to use. Many consumers nowadays are baffled by the multitude of push-buttons and switches that are available on their equipment, and very rarely touch more than one or two of the controls. Human-friendly design becomes more and more mandatory. This is what true sophistication stands for: to please the customer and make him feel at ease.

Summary. Text of a speech given in Nanjing in the People's Republic of China on 1st June 1988 by Dr S. van Houten, member of the Philips Board of Management. After a short introduction to the Philips company the speaker shows how important the role of scientific research is in bringing out electronic consumer products. He gives examples for a number of important fields such as lighting (economical lamps), television technology (various kinds of displays, high-definition television), magnetic and optical recording (Compact Cassette, Compact Disc) and integrated circuits (sub-micron technology). The speaker makes various points in conclusion, emphasizing that the traditional distinction between professional products and consumer products has changed its character completely and that the user-friendliness of consumer products will be a factor of growing significance.

^{19]} W. G. Gelling and F. Valster, The new centre for submicron IC Technology, Philips Tech. Rev. 42, 266-273, 1986.

Sound radiation from a vibrating membrane

J. H. Streng

'Mr. Watson, come here, I want you' are reputed to be the first intelligible words that Alexander Graham Bell spoke via his telephone connection. On 14th February 1876 he filed the first application for a patent on a telephone. Whether he should really be seen as the inventor seems to be a matter of some doubt⁽¹⁾. There is however no doubt that the development of electro-acoustic transducers, essential components in telephony, has flourished since that first patent application. There are various mechanisms for the conversion of electrical energy into acoustic energy in such transducers, such as the vibration of an electrically charged membrane in a varying electric field. This article deals with the radiation of sound from such a membrane. It is shown that the combination of classical analytical mathematics and modern computer techniques has led to a better understanding of this sound radiation and that there is an interesting application.

Introduction

In electroacoustic transducers such as telephones and loudspeakers electrical energy is converted into acoustic energy. Electrical quantities give rise to forces that in turn set up air vibrations. The conversion of electrical quantities into mechanical quantities is based on the interaction of magnetic fields and electric currents — Lorentz forces — or on the interaction of electric fields and charges — electrostatic forces. Electroacoustic transducers are therefore mainly classified as either electrodynamic or electrostatic versions. If we now just consider loudspeakers, we find that electrodynamic loudspeakers are by far the most widely used. In the most common type a coil energized by alternating current moves in a magnetic field produced by a permanent magnet. The forces acting on the coil cause a membrane — usually conical — to vibrate.

The vibrating membrane produces sound waves in the surrounding air.

An *electrostatic* loudspeaker consists in its simplest form of a capacitor with air as the dielectric. One of the two electrodes takes the form of a tightly stretched flat membrane, and the other is a stationary plate. When the electrodes are connected to the two poles of a high-voltage direct-current source, the electrodes receive charges of opposite sign. Each electrode is then subjected to a constant force equal to the product of field-strength and charge. If an alternating voltage is superimposed on the constant voltage, the constant forces acquire an alternating component. The result is that the membrane will vibrate at a frequency that as a first approximation is equal to the frequency of the alternating voltage. The vibrations of the membrane set up sound waves in the surrounding air. Later in this article it will be shown that the nonlinear behav-

Ir J. H. Streng is with Philips Research Laboratories, Eindhoven.

our of this simplest electrostatic loudspeaker can be improved if two stationary electrodes are used, instead of one, with the membrane between them.

In the conversion of electrical power into sound power a loudspeaker behaves as a highpass filter with a transfer characteristic that rises by a constant number of decibels per octave in the low-frequency range. The cut-off frequency of this 'filter' is inversely proportional to the diameter of the loudspeaker, whether electrodynamic or electrostatic. The loudspeaker behaves like a highpass filter because it cannot build up static pressure in the surrounding air. It follows from the shape of the filter response that a loudspeaker should have a large diameter and hence a large area if it is to produce sound at high power at low frequencies.

All loudspeakers tend to some extent to focus the sound in the direction of the principal axis^[1]. This focusing effect — which is usually undesirable — is due to interference between sound waves emitted by different parts of the vibrating membrane^[2]. Interference becomes perceptible when the dimensions of the membrane are of the same order of magnitude as the wavelength of the sound — about 30 cm at 1000 Hz. The effect increases as the dimensions become larger than the wavelength. To prevent the focusing of high-frequency sound the dimensions of the vibrating membrane should be small. This is one of the reasons why electrodynamic loudspeakers generally have separate loudspeakers for different frequency ranges, with filter circuits to divide the output signal from the amplifier into different signals for each range. Electrodynamic loudspeakers have the disadvantage, however, that the mechanical properties of the membrane have a considerable effect on the shape of the transfer characteristic.

Electrostatic loudspeakers have not been as widely used as electrodynamic loudspeakers, mainly because of mechanical and electrical limitations:

- High voltages have to be used (up to 5 kV or so, a.c. and d.c.).
- The membrane of an electrostatic loudspeaker cannot be deflected as much as that of an electrodynamic loudspeaker, because there is nonlinear distortion if the membrane is stretched too far.
- If the force acting on the membrane of an electrostatic loudspeaker is increased, the field-strength in the air gap between the electrodes may become so high — 2 kV/mm or more — that there is electrical breakdown.

To keep both the deflection and the field-strength within reasonable limits, protection circuits are necessary. These circuits can now be produced more cheaply by integrating them on silicon. The effect of

limiting the deflection of the membrane is that the area of an electrostatic loudspeaker that can reproduce the lower audio frequencies must be large, 0.5 m² or more. This is why electrostatic loudspeakers have so far only been used for the higher audio frequencies.

An investigation into the acoustic radiation from vibrating membranes has been in progress for some years at Philips Research Laboratories in Eindhoven. Earlier calculations of the sound field were based on the model of a rigid piston moving in air^[2] [3]. This model differs from that of a vibrating membrane, however, because a rigid piston has the same deflection at all points of its surface. Also, the edge effects are essentially different: the displacement of a membrane at its circumference is zero, unlike that of a piston. This means that there are turbulences at the edge of a piston that are not found at the edge of a vibrating membrane.

We have devised a method of solving the set of differential equations consisting of:

- an equilibrium equation for the vibrating membrane,
- Helmholtz's equation for the propagation of sound in the surrounding air, and
- an equation that takes into account an 'interface' condition.

The computer program developed for solving this set of equations calculates the membrane movement and the sound pressure on the membrane for arbitrary values of the mechanical and electrical parameters. It is therefore possible to determine the magnitude and direction of the sound intensity (i.e. the energy flux density in W/m²) at every point of a rectangular network in air; see *fig. 1*.

The investigation has resulted in a laboratory design for an electrostatic loudspeaker. The problem of high-frequency focusing due to the large area necessary for reproducing the low frequencies has been solved by subdividing the surface of the membrane. Signal components at higher frequencies are reproduced by a smaller part of the membrane. The signal is passed through a number of lowpass filters with decreasing cut-off frequencies. The output signal from each filter is applied to a separate output amplifier. The amplifier for the unfiltered signal is connected to the smallest part of the membrane. As the signal

[1] F. V. Hunt, *Electroacoustics*, Harvard University Press, Cambridge, Mass., 1954.

[2] The principal axis is defined as the centre-line of the conical membrane of an electrodynamic loudspeaker, or as the line through the centre and perpendicular to the flat membrane of an electrostatic loudspeaker.

[3] L. L. Beranek, *Acoustics*, McGraw-Hill, New York 1954.

[4] C. J. Bouwkamp, *Theoretische en numerieke behandeling van de buiging door een ronde opening*, (Theoretical and numerical treatment of the diffraction of a scalar plane wave through a circular aperture), Thesis, Groningen 1941.

passes through the chain of filters, the high-frequency components disappear, and the membrane areas connected to the filter outputs become larger.

This approach has given a loudspeaker with a useful frequency range from 50 to 20000 Hz. A particular feature is that the output amplifiers in the loudspeaker have optoelectronic couplers between the high-voltage and low-voltage sections. Another feature is that there is feedback to keep the non-driven part of the membrane at rest at frequencies above the filter cut-off frequencies. The optoelectronic couplers in the output amplifiers have the great advantage that the output current of the audio power amplifier can be lower than in conventional designs with isolating transformers. The output current is reduced by a factor equal to the turns ratio of the transformers. Another disadvantage of isolating transformers is that they are expensive.

In the rest of the article we shall first look at the set of three differential equations and consider ways of solving them. Some of the results of calculations made with our programs will then be presented. Finally, the electrostatic loudspeaker and its electronic drive will be described.

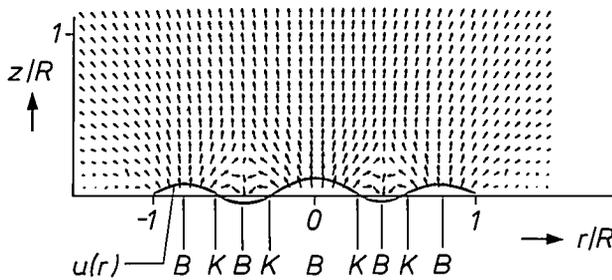


Fig. 1. Results of the calculation of the sound radiation from a vibrating circular membrane of maximum radius R . The arrows indicate the relative magnitude and direction of the sound intensity in the surrounding air as a function of the coordinates r and z , which have been normalized with respect to R . The derivatives with respect to the third coordinate θ are equal to zero. The blue line relates to the instantaneous deflection u of the membrane as a function of the radius r . There are two circular 'nodal lines' K where the membrane is approximately at rest. The nodal lines separate areas B where the membrane is vibrating.

Theoretical background

It should be sufficient here to state the three differential equations and briefly describe their solution. The method used has been described in detail elsewhere [4].

The complex amplitude η of the deflection as a function of the radius r of a circular membrane, see fig. 2, is one of the two unknowns in the differential

equation that describes the equilibrium of the perpendicular forces acting on the membrane per unit area:

$$T \nabla_2^2 \eta(r) + \omega^2 \rho_m \eta(r) - j\omega Z_s \eta(r) + F(r) - \{p_+(r) - p_-(r)\} = 0, \text{ for } r \leq R, \quad (1)$$

with the boundary condition $\eta(R) = 0$. For reasons of symmetry it is assumed that the various quantities are independent of the (tangential) angular coordinate θ . All terms in the equation are harmonic functions. The time-dependent factors $e^{j\omega t}$ have been omitted, leaving the complex amplitudes. First of all, the significance of the various terms will be explained.

The first term in eq. (1) represents the resultant force per unit area due to the forces acting on the edges of a circular surface element as a result of the tension in the membrane. T is the membrane tension in N/m, which has the same magnitude in all directions since we are dealing with a uniformly stretched membrane. The factor $\nabla_2^2 \eta(r)$ is the curvature of the membrane at the position r in cylindrical coordinates, written as:

$$\nabla_2^2 \eta(r) = \frac{\partial^2}{\partial r^2} \eta(r) + \frac{\partial}{r \partial r} \eta(r).$$

The second term in eq. (1) represents the force per unit area required to accelerate the membrane. In this term ρ_m is the density and ω is the angular velocity of the membrane vibrations.

The third term in eq. (1) takes account of acoustic damping that may be added by fabric or gauze. This extra damping, which is superimposed on the 'ordinary' damping caused by the surrounding air, is characterized by the specific acoustic impedance Z_s (i.e. the acoustic impedance for unit area). This impedance is defined as the complex ratio of pressure to air velocity.

The fourth term in equation (1) represents the force per unit area acting externally on the membrane.

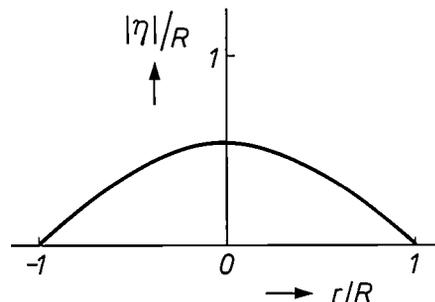


Fig. 2. The absolute value of the complex amplitude η of the membrane deflection as a function of the radius r , both normalized with respect to the maximum radius R .

The last two terms account for the pressure difference resulting from the displacement of air by the vibrating membrane. $p_+(r)$ is the sound pressure acting on the membrane from the positive z -direction, $p_-(r)$ is the sound pressure acting from the negative z -direction; see fig. 1. (The sound pressure is the difference between the instantaneous pressure and the mean pressure.) Because there is antisymmetry, we have: $p_-(r) = -p_+(r)$. The pressure $p_+(r) = -p_-(r)$ is the other of the two unknowns in differential equation (1).

The next differential equation is Helmholtz's equation^[2], which follows from the wave equation in three dimensions for a continuous medium, in this case air:

$$\nabla_3^2 p(r, z) + k^2 p(r, z) = 0, \quad (2)$$

with the boundary conditions: $p(r, 0^+) = p_+(r)$ and $p(r, 0^-) = p_-(r)$, both for $r \leq R$. The pressure p in free space is considered for reasons of symmetry to be a function of the coordinates r and z alone, and not of the third coordinate θ . The symbol 0^+ means that $z \rightarrow 0$ from the positive z -direction; similarly 0^- means that $z \rightarrow 0$ from the negative z -direction. The squared three-dimensional nabla operator in cylindrical coordinates can be written (since partial derivatives with respect to θ are zero, of course) as:

$$\nabla_3^2 p(r, z) = \frac{\partial^2}{\partial r^2} p(r, z) + \frac{\partial}{\partial r} p(r, z) + \frac{\partial^2}{\partial z^2} p(r, z).$$

The wave number k is equal to ω/c_0 , where c_0 is the velocity of sound, and also equal to $2\pi/\lambda$, where λ is the wavelength.

In solving the differential equations (1) and (2) an interface condition must also be taken into account. This equates the acceleration of the air at the surface of the membrane to the acceleration of the membrane:

$$\left. \frac{\partial p(r, z)}{\partial z} \right|_{z=0^+} = \left. \frac{\partial p(r, z)}{\partial z} \right|_{z=0^-} = \omega^2 \rho_0 \eta(r), \quad \text{for } r \leq R. \quad (3)$$

In this third differential equation ρ_0 is the density of air.

Now a solution has to be found for the set of three differential equations (1), (2) and (3). The method we have adopted makes use of Green's functions^[5]. Let us first consider the complex amplitude η of the deflection of the membrane at the radius r . The deflection is partly due to pressures of complex amplitude $p_+(r)$ and $p_-(r)$, acting on opposite sides of the membrane. Now we consider the pressures at different radii r_0 as

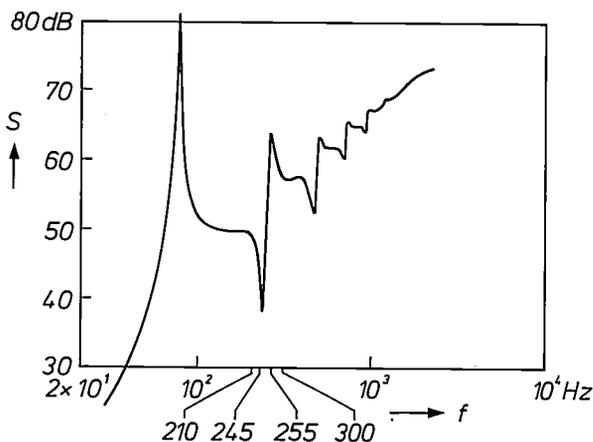


Fig. 3. The calculated sound pressure level S in dB^[6] at the principal axis as a function of the frequency f on a logarithmic scale. The sound pressure level is calculated for a distance of 1 m from the surface of a circular membrane. The values of the parameters are: membrane tension $T = 100$ N/m, density of the membrane $\rho_m = 0.02$ kg/m², maximum radius $R = 0.125$ m, acoustic impedance $Z_s = 0$, force per unit area $F = 1$ N/m².

'sources', each contributing to the deflection at r . Then the effect of all these sources from $r_0 = 0$ to $r_0 = R$ can be summed in such a way that the result gives the total deflection at r . It is a property of the Green's function $G_1(r|r_0)$, which is a function of the two variables r and r_0 , that a factor is added to the separate sources such that the summation gives exactly the deflection:

$$\eta(r) = \int_0^R p_+(r_0) G_1(r|r_0) r_0 dr_0. \quad (4)$$

In much the same way the pressures $p_+(r)$ and $p_-(r)$ on opposite sides of the membrane at r_0 can be regarded as sources that contribute to the sound pressure $p(r, z)$ in free space. The Green's function $G_2(r, z|r_0, z_0)$ adds a factor to the individual sources such that the total sound pressure at r and z is obtained. The result is the integral equation:

$$p(r, z) = \int_0^R p_+(r_0) \left\{ \frac{\partial}{\partial z_0} G_2(r, z|r_0, z_0) \right\} \Big|_{z_0=0} r_0 dr_0. \quad (5)$$

If we substitute (5) in equation (3) we obtain a second

[4] J. H. Streng, Calculation of the surface pressure on a vibrating circular stretched membrane in free space, J. Acoust. Soc. Am. 82, 679-686, 1987;
 J. H. Streng, Calculation of integrals which occur in circular stretched membrane sound radiation, J. Acoust. Soc. Am. 83, 1183-1185, 1988;
 J. H. Streng, Sound radiation from circular stretched membranes in free space, Proc. 84th AES Conv., preprint 2573 (C-2), 1-41, 1988.
 [5] P. M. Morse and K. U. Ingard, Theoretical acoustics, McGraw-Hill, New York 1968.

integral equation for $\eta(r)$:

$$\eta(r) = \frac{1}{\omega^2 \rho_0} \int_0^R p_+(r_0) \left\{ \frac{\partial}{\partial z} \left(\frac{\partial}{\partial z_0} G_2(r, z | r_0, z_0) \right) \right\} \Big|_{z=0} \Big|_{z=0^+, 0^-} r_0 dr_0. \quad (6)$$

If we now equate $\eta(r)$ as given by eq. (4) with $\eta(r)$ as given by eq. (6), we obtain an integral equation in which the only unknown is the pressure p_+ on the surface of the membrane. From this we can solve for p_+ , by expressing p_+ in terms of a power series whose n coefficients can be determined by solving n equations in n unknowns. Once p_+ is known, we can calculate the complex amplitudes $\eta(r)$ for the membrane displacement and $p(r, z)$ for the sound pressure using equations (4) and (5).

Results of calculations

Our program can provide the following results:

- the frequency spectrum of the sound pressure level S in dB [6] at any distance from a vibrating circular membrane;
- the direction and magnitude of the sound intensity in the air, represented as vectors at the nodes of a rectangular network;
- the absolute value of the complex amplitude p_+ of the sound pressure on the membrane surface as a function of the radius r for a given frequency, with the corresponding phase angle ϕ with respect to the driving force (F in eq. 1);
- the absolute value of the complex amplitude η of the membrane deflection, also as a function of r and with the corresponding phase angle ϕ .

The parameters that can be set as required for each calculation cycle are the membrane tension T , the density ρ_m and the maximum radius R of the membrane, the specific acoustic impedance Z_s , which characterizes the additional damping of the membrane, and the amplitude F of the force driving the membrane.

Fig. 3 shows a frequency spectrum for the sound pressure level on the axis at a distance of 1 m from a membrane with a diameter of 25 cm. The spectrum extends from 30 to 2000 Hz and applies in the case in which the movement of the membrane does not have extra damping ($Z_s = 0$, the values for the other parameters are given in the caption). At low frequencies the sound pressure level S increases by 18 dB per octave. It can be shown that this means that the real part of the acoustic impedance of the air is proportional to ω^4 , which in turn shows that the vibrational behaviour of the membrane is very like that of a rigid piston, with the entire membrane vibrating in phase [2].

At 75 Hz there is a sharp resonance peak, which shows that at this frequency the mass of the air load and the membrane tension form a mass-spring system.

At frequencies above 1000 Hz S increases by 6 dB per octave on average. In this frequency range the membrane behaves in such a way that a succession of annular areas, which theoretically should vibrate in opposite phase, move radially across the membrane [4]. These 'antinodes' are separated by circular 'nodal lines'; see fig. 1. The average increase of S by 6 dB per octave indicates an acoustic impedance of the air independent of ω . For a plane-wave source this impedance is the product of the sound velocity and the density of air, so that it is indeed independent of ω [2].

Because of the damping action of the air the membrane is not exactly stationary at the nodal lines, so that the nodal lines are only apparent. As the frequency increases, the actual behaviour of the membrane in air departs more and more from the theoretical behaviour of the membrane 'in vacuo'. At high frequencies the membrane therefore becomes very like a plane-wave source. This is less so at low frequencies because the vibration of the individual areas in opposite phase produces 'acoustic short-circuiting'.

The behaviour of the membrane in the low-frequency and high-frequency ranges is not surprising. The increase in S at high frequencies mentioned earlier is due to the focusing of the sound along the axis. What is more interesting, however, is the behaviour of S in the range from 100 to 1000 Hz, where a succession of similar resonance effects appears.

We have investigated this frequency range in more detail by studying the membrane deflection, the sound pressure at the surface and the sound-intensity distribution at the frequencies 210, 245, 255 and 300 Hz; see fig. 4. We see that at 210 Hz the surface pressure and the membrane deflection show an apparent nodal line in their amplitude, and that the pressure in the area inside the nodal line is in phase with the driving force, while the deflection in this area is out of phase with it. As the frequency increases to 245 Hz, both the nodal lines move radially inwards. The nodal pressure

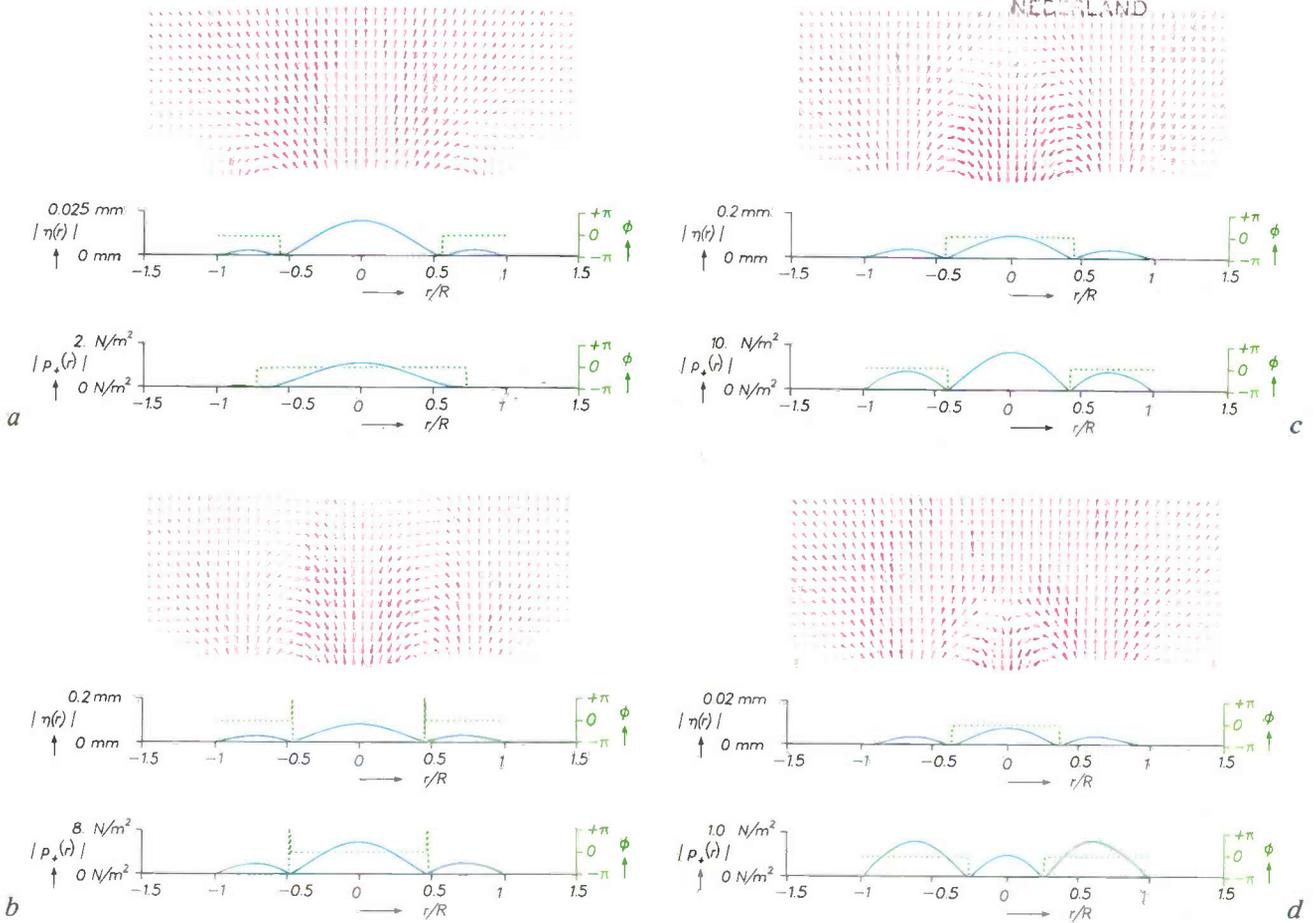


Fig. 4. Results of calculations for the frequencies a) 210, b) 245, c) 255 and d) 300 Hz. The absolute values of the complex amplitudes p_+ of the sound pressure at the surface and η of the membrane deflection are plotted as a function of the radius r (blue lines). The sound intensity distribution (red arrows) and the phase angle ϕ relative to the force F (green dashed lines) are also shown. The results apply for the same parameters as in fig. 3. See also the caption to fig. 1.

line moves inwards faster than the nodal deflection line and eventually passes it. The membrane behaves like a combination of two mass-spring systems: the frequency spectrum (fig. 3) has a deep ‘antiresonance’ dip that is followed by a sharp resonance peak. At 255 Hz the peak in the spectrum has been passed, as can be seen from the 180° phase jump in both pressure and deflection, both inside the nodal line and outside it. It must be concluded that the frequency at which the nodal line for the pressure passes that for the deflection corresponds to the succession of an antiresonance and resonance in the frequency spectrum.

At 300 Hz the situation is again comparable with that at 210 Hz. As the frequency increases, a new set of nodal lines for pressure and deflection will appear at the position $r = R$. These nodal lines also pass each

other and again produce a dip and a peak in the spectrum; see fig. 3. In this way more and more nodal lines appear, which separate areas vibrating in opposite phase. At higher frequencies the nodal lines become less pronounced owing to the greater damping effect of the air. Above 1000 Hz the acoustic impedance becomes independent of frequency, as we saw earlier. At the higher frequencies the acoustic impedance is no longer reactive or capacitive but purely resistive.

It can also be seen from fig. 5 that the sound is focused along the principal axis at high frequencies. This figure shows the sound spectrum obtained when

[6] The SPL (‘sound pressure level’) in dB is defined as 20 times the logarithm to the base 10 of the ratio of the r.m.s. value of the sound pressure to a standardized reference pressure of 2×10^{-5} N/m², which corresponds approximately to the threshold of human hearing.

the sound pressure level is calculated on a line at 45° to the axis, again at a distance of 1 m from the source. It is evident that the sound-pressure level decreases with frequency above 1000 Hz.

In most applications strong resonances of the membrane, like those in figs 3 and 4, are undesirable. We have therefore calculated the frequency spectrum for the case in which the specific acoustic impedance Z_s for the additional damping of the membrane differs from zero. Fig. 6 shows three frequency spectra, for $Z_s = 0, 20$ and 40 Ns/m^3 . At the highest damping the peaks and dips in the spectrum are almost completely

one of the electrodes. More sound radiation therefore implies a higher membrane tension in an electrostatic loudspeaker. We shall leave these effects here, and in the next section it will be shown how the better understanding we had obtained led to a practical application.

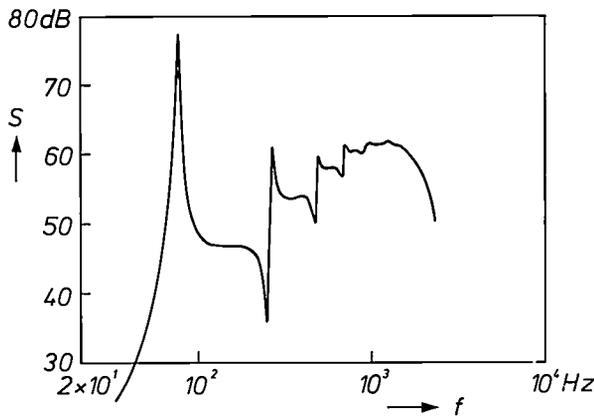


Fig. 5. Sound pressure level S as a function of frequency f for the same parameters as in fig. 3, but now at an angle of 45° to the principal axis.

suppressed. Additional damping can be introduced by fitting an 'acoustically transparent' piece of fabric parallel and fairly close to the membrane. A damping of 40 Ns/m^3 corresponds to the damping caused by a piece of thin cotton fabric.

We varied other parameters as well, such as the mass and the tension of the membrane. As expected, increasing the mass and reducing the tension moves the resonance peaks and dips to lower frequencies. We also examined the effect of a negative compliance, i.e. the effect of a force on the membrane with the same sign as the deflection, and increasing with the deflection. The voltage applied to the electrodes of an electrostatic loudspeaker produces a negative compliance, for example. It has been found that there is instability when the effect of this compliance is greater than that of the membrane tension. If we increase the voltage on the electrodes, we must also increase the tension of the membrane, otherwise the membrane will 'break out' and attach itself electrostatically to

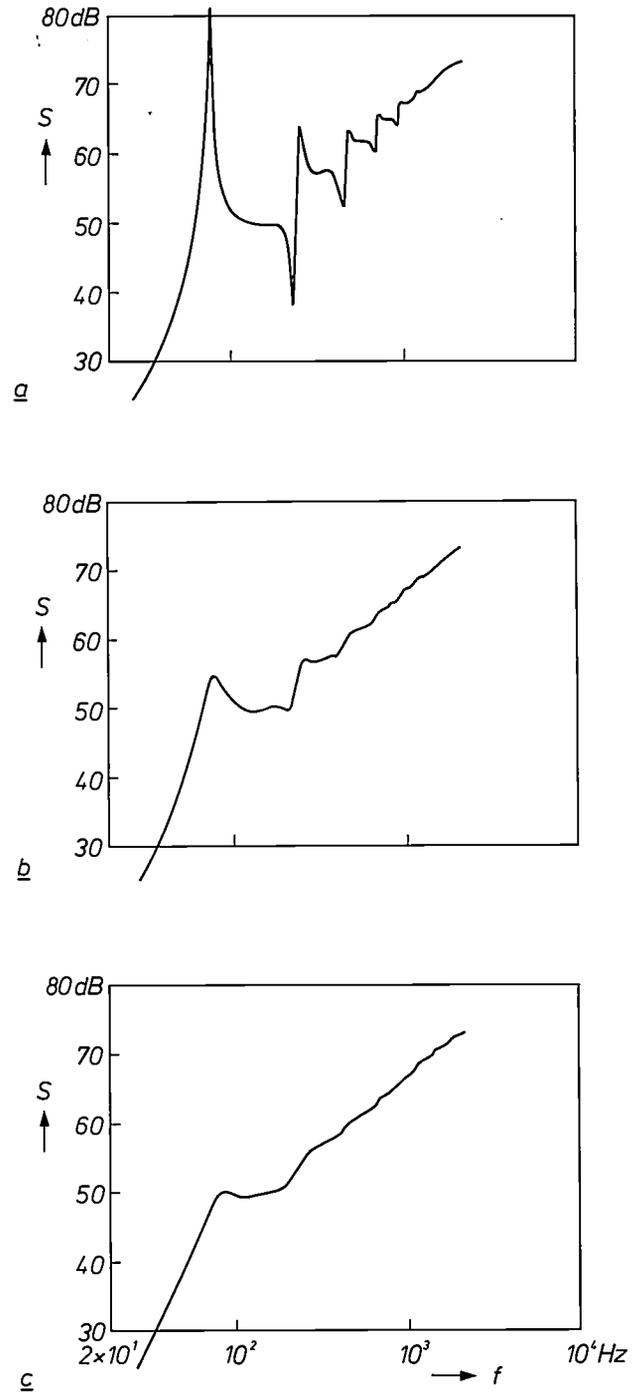


Fig. 6. Sound pressure level as a function of frequency, a) for the parameters in fig. 3, b) for the same parameters but with acoustic impedance Z_s of 20 Ns/m^3 and c) as before but with $Z_s = 40 \text{ Ns/m}^3$.

An electrostatic loudspeaker

In the introduction the principle of a single electrostatic loudspeaker was described very broadly: the loudspeaker is a stretched membrane acting as one of the electrodes of a capacitor. Most electrostatic loudspeakers, however, are designed to operate on the 'push-pull' principle; see *fig. 7*. The membrane is located between two stationary electrodes, which are perforated with a number of small holes to allow the passage of sound waves.

The push-pull electrostatic loudspeaker operates as follows. A d.c. source supplies a high voltage V_0 . One terminal of this source is connected to the membrane through a high resistance R_0 and the other is connected to the two stationary electrodes. The source provides the charge for both capacitors. The audio signal is represented by the two a.c. sources V_1 . The

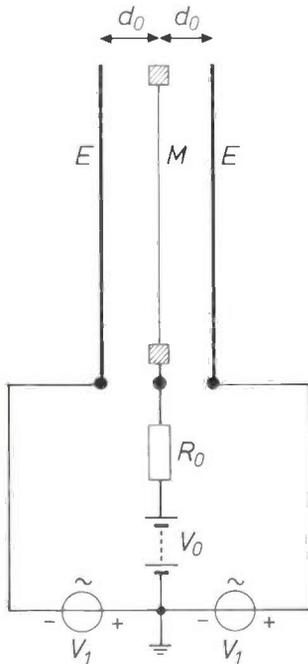


Fig. 7. Schematic illustration of a push-pull electrostatic loudspeaker. M stretched membrane. E stationary electrodes. R_0 high resistance. V_0 high-voltage source. V_1 a.c. voltage sources. d_0 electrode spacing.

alternating voltages give rise to alternating forces F per unit area on the charges ^[7]:

$$F = 2 \epsilon \frac{V_0 V_1}{d_0^2}, \quad (7)$$

where d_0 is the spacing of the plates and ϵ is the permittivity of air. The alternating force F makes the membrane vibrate.

In one of our laboratory designs for an electrostatic speaker the area of the stretched membrane is about 0.5 m^2 and the maximum deflection is 3 mm. To produce this deflection it is necessary to give V_0 a con-

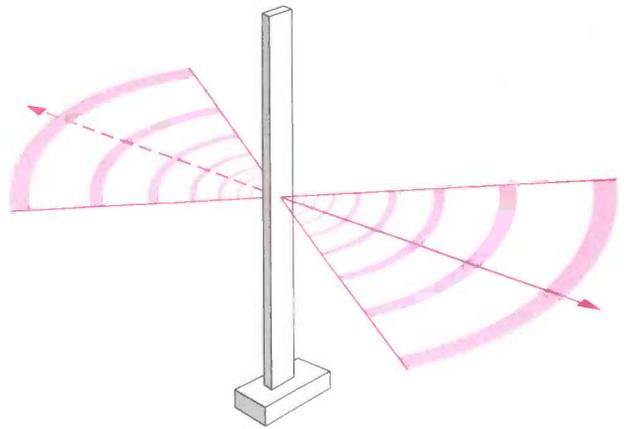
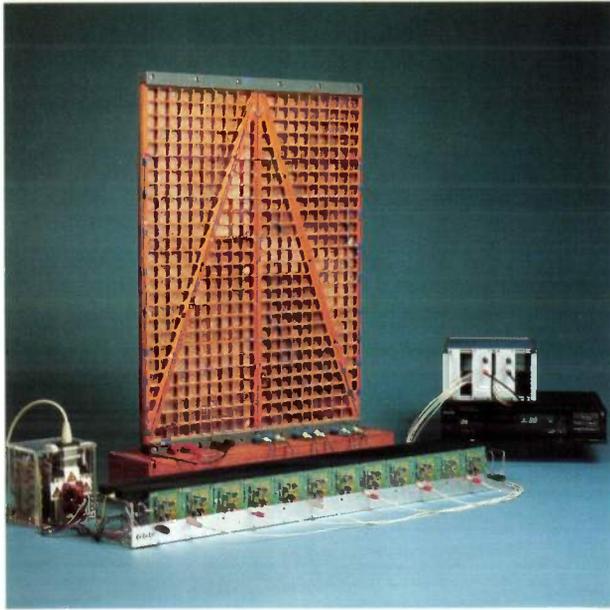


Fig. 8. Electrostatic column loudspeaker. The sound is focused in the vertical plane and there are difficulties because of the 'comb' effect.

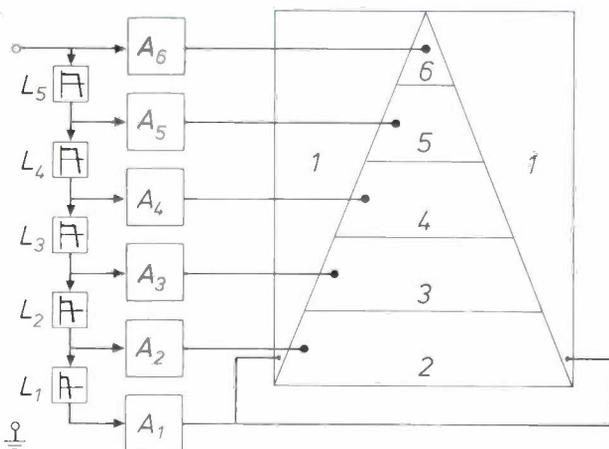
stant value of 4000 V and V_1 a maximum amplitude of 4000 V. Obviously, this places some rather special requirements on the material of the electrodes. The membrane consists of an extremely thin polyimide film, coated on one side with graphite. The graphite has a fairly high resistance. The charge cannot therefore move over the surface of the capacitor easily, which might make the field-strength rise locally to the breakdown value. The stationary electrodes are perforated, as we saw just now, and consist of epoxy resin reinforced with glass fibre and coated with copper on the outside. The material is also used for printed circuits. The epoxy resin thus acts as insulation for the actual electrodes. An electrostatic loudspeaker, unlike an electrodynamic loudspeaker in a cabinet, also radiates sound backwards.

To give good reproduction at lower frequencies the electrostatic loudspeaker has a large radiating area. However, to suppress the focusing effect at higher frequencies the dimensions should be small — much smaller than the wavelength of the sound. There are loudspeaker manufacturers who have solved this problem by designing the loudspeaker as a narrow but very high column, again with a radiating area of about 0.5 m^2 . In these loudspeakers the higher-frequency sound is only focused in the vertical plane, and fans out in a horizontal plane; see *fig. 8*. The objections to the column shape are that the loudspeaker

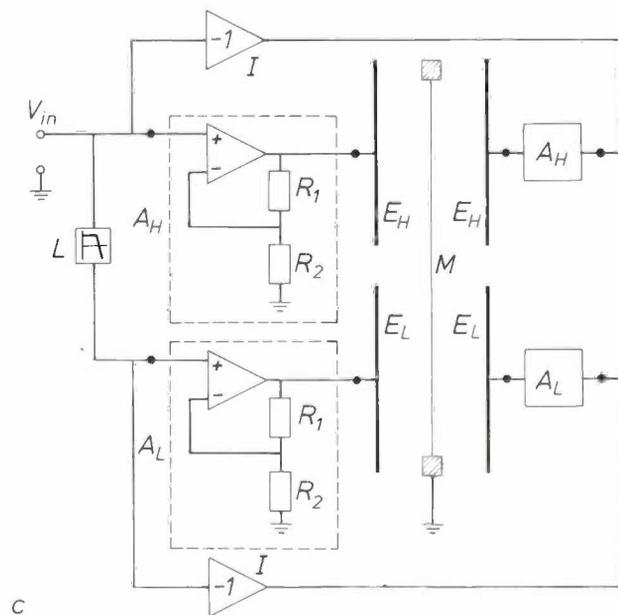
[7] F. V. Hunt, *Electroacoustics*, Harvard University Press, Cambridge, Mass., 1954.



a



b



c

Fig. 9. a) Laboratory design for an electrostatic loudspeaker with a total radiating area of 0.5 m^2 . b) Subdivision of the surface into areas 1 to 6. A_1 to A_6 electronic circuits. L_1 to L_5 lowpass filters with increasing cut-off frequency. c) Circuit for the simplified case in which the loudspeaker area is subdivided into two parts, denoted by subscripts H and L. V_{in} input voltage, the output voltage of an audio power amplifier. L lowpass filter. A_H and A_L circuits each consisting of an amplifier, an optoelectronic coupler (not shown) and a voltage divider, resistances R_1 and R_2 . The voltage dividers give high feedback of the output voltage from the amplifiers. E_H and E_L stationary electrodes. M membrane. I amplifiers of gain -1 , which drive the stationary electrodes in opposite phase on opposite sides.

<

is impractically high, 2.5 m perhaps, and the higher frequencies are only perceived in a horizontal region of extremely limited height. Also, this horizontal region contains zones in which the sound is cancelled out — the comb effect — because of destructive interference between sound originating from the centre of the loudspeaker and sound from the extremities.

Our answer to these problems is to reduce the effective radiating area of the loudspeaker as the frequency increases. We subdivided the area of the loudspeaker as shown in fig. 9. At the low frequencies all the areas 1 to 6 are operative. The lowpass filters L_1 to L_5 , whose cut-off frequencies increase with increasing subscript number, ensure that the circuits A_1 to A_6 receive signals with an increasing frequency range. The two areas 1 are driven by the lowest-frequency components of the audio signal; area 6 is driven by the complete audio signal. Area 6 is so small that there is no undesirable focusing even at the high frequencies.

A problem in subdividing the membrane area as shown in fig. 9b is to prevent the high-frequency vibrations from 6, say, from causing the surrounding areas of the membrane to vibrate. Fig. 9c shows how we solved this problem. (The figure is based on a simplified subdivision of the membrane into two parts; the corresponding components are given the subscripts H and L.) Each part of the loudspeaker is driven by two circuits that each include a high-voltage amplifier with high feedback. The two parts of the stationary electrodes on either side are driven by these two circuits in opposite phase with the aid of an amplifier of gain -1 .

The special feature here is that the high feedback in the high-voltage amplifiers effectively switches off the parts of the loudspeaker that do not receive any signal. This is because each part of the loudspeaker is a reciprocal element in which a deflection of the membrane induces a signal even though no input signal is present. The feedback then causes a negative output

[8] The electronic couplers were developed by P. A. Dessens and K. Oostveen of Philips Research Laboratories, Eindhoven.

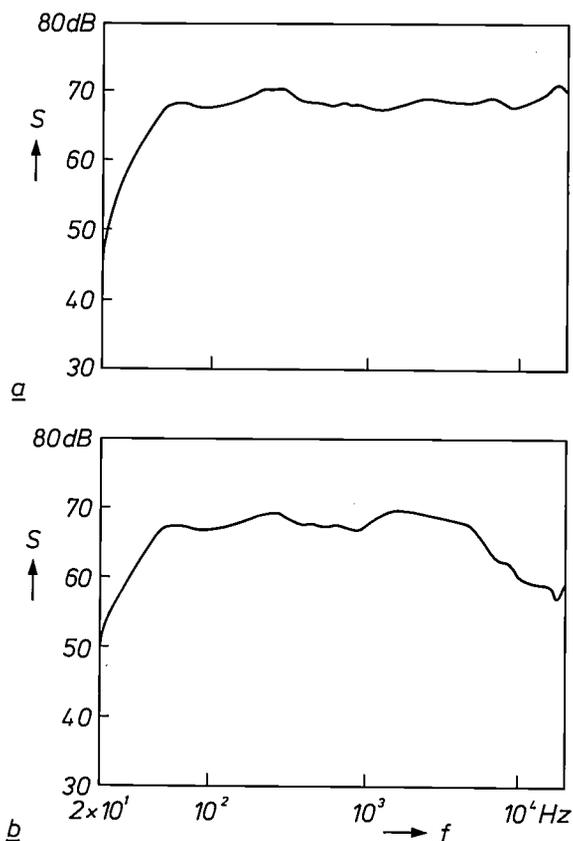


Fig. 10. Characteristics of the first laboratory version of the electrostatic loudspeaker of fig. 9. *a*) Sound pressure level S in dB^[6] at a distance of 2 m on the principal axis as a function of frequency f . *b*) Similar characteristic, but at an angle of 20° to the axis. The curves (*a*) and (*b*) are almost identical. Only at frequencies above 7 kHz is curve (*b*) perceptibly below curve (*a*). This indicates that the focusing effect in the electrostatic loudspeaker at high frequencies has been greatly suppressed. Later versions will have even better characteristics.

voltage to appear at the output of the amplifier, which suppresses the original deflection.

Another feature is that we do not use high-voltage transformers as in the electrostatic loudspeakers now on the market. High-voltage transformers are expensive, because they must be highly linear and handle high powers. In our circuits we use a particular kind of optoelectronic coupler (not shown in fig. 9c)^[8]. The output stage of the amplifier in this arrangement supplies the high-voltage signal and consists of a circuit containing photosensitive transistors. The output of the power amplifier stage is formed by light-emitting diodes (LEDs). The light emitted by the LEDs drives the transistors of the output stage. This means that there is no electrical connection between the output stage and the input of the amplifier, except for the feedback via the voltage divider (fig. 9c). Finally, fig. 10 shows the characteristics of the first laboratory version of the new electrostatic loudspeaker.

Summary. The sound radiation from a vibrating membrane can be described by three differential equations: the equation of motion, Helmholtz's equation and an interface condition. This set of equations can be solved with the aid of Green's functions. The results, presented by means of a computer program, show that annular nodes and antinodes occur on a circular stretched membrane, which increase in number with rising frequency. The improved theoretical understanding obtained has led to a laboratory design for an electrostatic loudspeaker. The problem of axial focusing of the higher frequencies has been solved by subdividing the surface of the membrane into separate active areas, each driven by a separate amplifier with a corresponding lowpass filter. Optoelectronic couplers are used in the amplifiers to avoid the need for high-voltage transformers.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale, P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

M. P. J. G. Versleijen, P. I. Kuindersma, G.-D. Khoe & L. J. Meuleman	E	Accurate analysis of dc electrical characteristics of 1.3µm DCPBH laser diodes	IEEE J. QE-23	925-935	1987
J. H. Streng	E	Calculation of integrals which occur in the analysis of circular stretched membrane sound radiation	J. Acoust. Soc. Am. 83	1183-1185	1988
C. Colinet*, A. Pasturel* (*Lab. Thermodyn. & Physico-Chimie Métallurgiques, St. Martin d'Hères) & K. H. J. Buschow	E	Short-range order and stability in Gd-Ni and Y-Ni systems	J. Appl. Phys. 62	3712-3717	1987
J. Haisma, A. M. W. Cox, B. H. Koek, D. Mateika, J. A. Pastorius & E. T. J. M. Smeets	E, H	Heteroepitaxial growth of InP on garnet	J. Cryst. Growth 87	180-184	1988
R. Grössinger*, R. Krewenka*, H. R. Kirchmayr* (*Univ. of Technol., Vienna), S. Sinnema ⁺ , Y. Fu-Ming ⁺ , H. Ying-Kai ⁺ , F. R. de Boer ⁺ (*Univ. Amsterdam) & K. H. J. Buschow	E	Magnetic anisotropy in Pr ₂ (Fe _{1-x} Co _x) ₁₄ B compounds	J. Less-Common Met. 132	265-272	1987
R. B. Helmholtz*, J. J. M. Vlegaar* (*ECN, Petten) & K. H. J. Buschow	E	Note on the crystallographic and magnetic structure of YFe ₁₀ V ₂	J. Less-Common Met. 138	L11-L14	1988
K. H. J. Buschow, P. Schobinger-Papamantellos (Inst. für Kristallogr. & Petrogr., Zürich) & P. Fischer (Lab. für Neutronenstreuung, Würenlingen)	E	Magnetic structure and properties of equiatomic rare earth germanides	J. Less-Common Met. 139	221-231	1988
C. J. M. Denissen, B. D. de Mooij & K. H. J. Buschow	E	Structure and ⁵⁷ Fe Mössbauer effect in R ₂ Fe ₁₄ C compounds	ibid.	291-298	1988
C. J. van der Poel	E	Rapid crystallization of thin solid films	J. Mater. Res. 3	126-132	1988
J. C. M. Henning, J. P. M. Ansems & P. J. Roksnoer	E	A photoluminescence study of the donor structure in Al _x Ga _{1-x} As	Semicond. Sci. Technol. 3	361-364	1988
M. M. Abd-Elmeguid*, B. Schleede*, H. Micklitz* (*Univ. Bochum), T. T. Palstra ⁺ , G. J. Nieuwenhuys ⁺ (*Univ. Leiden) & K. H. J. Buschow	E	The stability of the ferromagnetic state in La(Fe _{0.86} Al _{0.14}) ₁₃ under high pressure	Solid State Commun. 63	177-180	1987
H. Verweij	E	Defect chemistry of YBa ₂ Cu ₃ O _x : strong evidence for the presence of Cu ⁺	Solid State Commun. 64	1213-1216	1987
P. Deppe*, M. Rosenberg* (*Univ. Bochum) & K. H. J. Buschow	E	A Mössbauer spectroscopy study of Nd ₅ Fe ₂ B ₆	ibid.	1247-1251	1987



M. G. J. Heijman, J. H. W. Kuntzel and G. H. J. Somers,
Multi-track magnetic heads in thin-film technology,
 PhilipsTech.Rev.44, No.6, 169-178, Dec. 1988.

CLS loggers (CLS stands for Communication-Logging System) used for logging spoken messages have to record or play back a large number of signals simultaneously at a very low tape speed. Multiple magnetic heads have to be used. Thin-film technology is used for fabricating 32-track recording heads that are different from the 32-track playback heads. The recording heads have the conventional coils, the playback heads have a magnetoresistance element. In the magnetoresistance element the resistance of a permalloy strip changes when the field in the magnetic 'yoke' of the playback head changes. The 32-track recording and playback heads are fabricated on a ferrite substrate by process steps like those used in IC technology: sputtering, electrochemical growth, photoresist deposition, exposure, plasma etching, ion etching, and photoresist removal. Test conductors in the heads enable tests to be carried out on the wafer. After all the tests have been carried out the multiple heads that have passed are sawn from the wafer and assembled to form a unit ready to be mounted in a logger.

J. H. Streng, Sound radiation from a vibrating membrane,
 PhilipsTech.Rev.44, No.6, 190-199, Dec. 1988.

The sound radiation from a vibrating membrane can be described by three differential equations: the equation of motion, Helmholtz's equation and an interface condition. This set of equations can be solved with the aid of Green's functions. The results, presented by means of a computer program, show that annular nodes and antinodes occur on a circular stretched membrane, which increase in number with rising frequency. The improved theoretical understanding obtained has led to a laboratory design for an electrostatic loudspeaker. The problem of axial focusing of the higher frequencies has been solved by subdividing the surface of the membrane into separate active areas, each driven by a separate amplifier with a corresponding lowpass filter. Optoelectronic couplers are used in the amplifiers to avoid the need for high-voltage transformers.

S. van Houten, Applied research — the source of innovation in consumer electronics,
 PhilipsTech.Rev.44, No.6, 180-189, Dec. 1988.

Text of a speech given in Nanjing in the People's Republic of China on 1st June 1988 by Dr S. van Houten, member of the Philips Board of Management. After a short introduction to the Philips company the speaker shows how important the role of scientific research is in bringing out electronic consumer products. He gives examples for a number of important fields such as lighting (economical lamps), television technology (various kinds of displays, high-definition television), magnetic and optical recording (Compact Cassette, Compact Disc) and integrated circuits (submicron technology). The speaker makes various points in conclusion, emphasizing that the traditional distinction between professional products and consumer products has changed its character completely and that the user-friendliness of consumer products will be a factor of growing significance.

I wish to subscribe to

PHILIPS TECHNICAL REVIEW

(date)

(signature)

Please tick the appropriate box

Regular subscription 80 guilders or U.S. \$ 35.00 per volume

The subscription includes postage and will start with Vol. 44, No. 1.
 Please pay when you receive our invoice.

Student's subscription 32 guilders or U.S. \$ 14.00 per volume

Please send a copy of your student's card or other written proof that you are a student; valid for two volumes.

Name

Initials

Title

Address

stamp

as

postcard

**Administration Department
Philips Technical Review**

**Philips Research Laboratories
Building WY 136**

P.O. Box 80 000

**5600 JA Eindhoven
The Netherlands**

Contents

- Interference filters in projection television tubes** 201
L. Vriens, J. A. Clarke and J. H. M. Spruit
*Making more light incident within the aperture angle
of the lens system improves the brightness
of the projection screen everywhere*
- Then and Now (1939-1989)** 211
- Analysis of the injection-moulding process** 212
J. F. Dijkman
*Materials scientist, product designer, mould designer and rheologist
all see the task from different viewpoints
as they create a combined product*
- A true silicon compiler for the design of complex ICs
for digital signal processing** 218
J. L. van Meerbergen and H. De Man
*One of the keys to the exploitation of the full potential
of IC technology is the availability of software
for highly automated chip design*
- Scientific publications** 231

Interference filters in projection television tubes

L. Vriens, J. A. Clarke and J. H. M. Spruit

Just looking at a television is no longer sufficient. Viewers want to be involved in the events on the screen. To provide this sensation of 'being there', sets with larger screens, stereo sound and higher-quality images are being developed. Projection television offers the possibility of larger screens. In one of the latest improvements in projection television, interference filters are used in the cathode-ray tube.

Introduction

Philips have been doing research on television receivers for more than fifty years. From the beginning (1936/1937) there have been direct-view sets and projection sets^{[1][2]}. In both systems the picture is usually obtained from a cathode-ray tube (CRT). In a direct-view set the viewer looks directly at the picture on the faceplate of the tube. In a projection set the picture on the faceplate of a small tube is projected on to a large screen by an optical system.

At first it was thought that large television pictures would have to be obtained by using projection sets. This is because there is a vacuum inside the CRT envelope, and glass was not considered strong enough to withstand the forces that the external atmospheric pressure would produce on a large envelope. (In those days 'large' referred to a screen diagonal of about 40 cm.) A great deal of research therefore went into projection television^{[2][3]}. However, the problems with the strength of large glass envelopes were solved and this, together with the higher picture quality of direct-view television, made the application of projection television in the home unnecessary. Projection television remained attractive for professional appli-

cations where very large screens are required for large audiences, and projection sets were sold in small numbers for professional use.

After the introduction of colour television, more than thirty years ago, new problems were encountered with projection television. To obtain sufficiently bright projected colour pictures it is necessary to use separate 'red', 'green' and 'blue' projection tubes. Magnified images from these three tubes are superimposed on a single projection screen. The problems arise in getting these three images into register^[4] and avoiding colour artefacts on the projection screen.

In recent years, and particularly in the United States, there has been a growing demand in the consumer market for large-screen television sets with screen diagonals of 90 to 150 cm. Although direct-view sets with diagonals of up to 110 cm have been made, they have the disadvantage that the weight increases with

[1] J. van der Mark, An experimental television transmitter and receiver, Philips Tech. Rev. 1, 16-21, 1936, and by the same author, Television, Philips Tech. Rev. 1, 321-326, 1936.

[2] M. Wolf, The enlarged projection of television pictures, Philips Tech. Rev. 2, 249-253, 1937.

[3] P. M. van Alphen and H. Rinia, Projection-television receiver, I. The optical system for the projection, Philips Tech. Rev. 10, 69-78, 1948.

[4] T. Poorter and F. W. de Vrijer, The projection of colour-television pictures, Philips Tech. Rev. 19, 338-355, 1957/58.

the third power of the diagonal and the depth increases linearly with the diagonal. A direct-view set with a diagonal of 100 cm weighs more than 100 kg and will hardly go through an ordinary door in the home. Projection sets weigh about half as much as a direct-view set with the same diagonal. Moreover, picture quality in projection television has increased, as a result of a number of technical improvements. This has led to a renewed interest in projection television. In the near future, the introduction of high-definition television (HDTV) with larger picture formats will be a further stimulus to the application of projection television.

The technical improvements in projection television relate to all the major components:

- *Projection screens* have been made with a wide horizontal viewing angle (120 to 170°) and a small vertical viewing angle (10 to 20°), resulting in a brightness gain of about five times in a direction perpendicular to the screen. The screen consists of two elements, a Fresnel lens and a second element consisting of vertical lenticular structures, transparent light-scattering particles (called diffusors) and, on the outer face, black regions that absorb ambient light. The lenticular structures and the diffusors ensure that hardly any colour artefacts appear^[5].
- The development of *lenses* consisting of plastic elements with some aspherical surfaces^[6] has led to light-weight high-aperture projection lenses that give a brighter image with good resolution, while, more recently, hybrid lenses consisting of a mixture of glass and plastic elements have resulted in better thermal stability^[7]. A liquid between the faceplate of the tube and the first lens element cools the faceplate and provides optical coupling with less reflection, giving higher contrast in the projected picture.
- The *phosphors* used in the tubes also have been improved, to give a higher luminous efficacy and a longer life^[8].
- The *electron guns* have been improved as well, so that it is now possible to use the tubes at higher peak currents, without significant deterioration in the resolution^[9]. The higher peak currents improve the brightness of the image.
- Finally, there have been advances in *electronic control* and the three images can be projected in more accurate register.

In this article we describe a new development in projection-television tubes. There were two stages in this development:

- 1) the deposition of an interference filter^[10] on the inside of the tube between the phosphor layer and the faceplate, and
- 2) the application of a curved faceplate^[11].

The combination of these two improvements gives:

- a considerable gain in brightness on the projection screen,
- better colour rendering (chromaticity),
- improved resolution and contrast.

We shall look in turn at several aspects of the optical imaging, the operation of the interference filter and its application on a curved faceplate.

Principles of projection television

Fig. 1 shows a diagram of a rear-projection set containing the following components: a projection-television tube, a lens system which projects the picture on the faceplate of the tube on to the screen, a mirror that folds the light path, and a viewing screen, on which the image appears. In practice the system con-

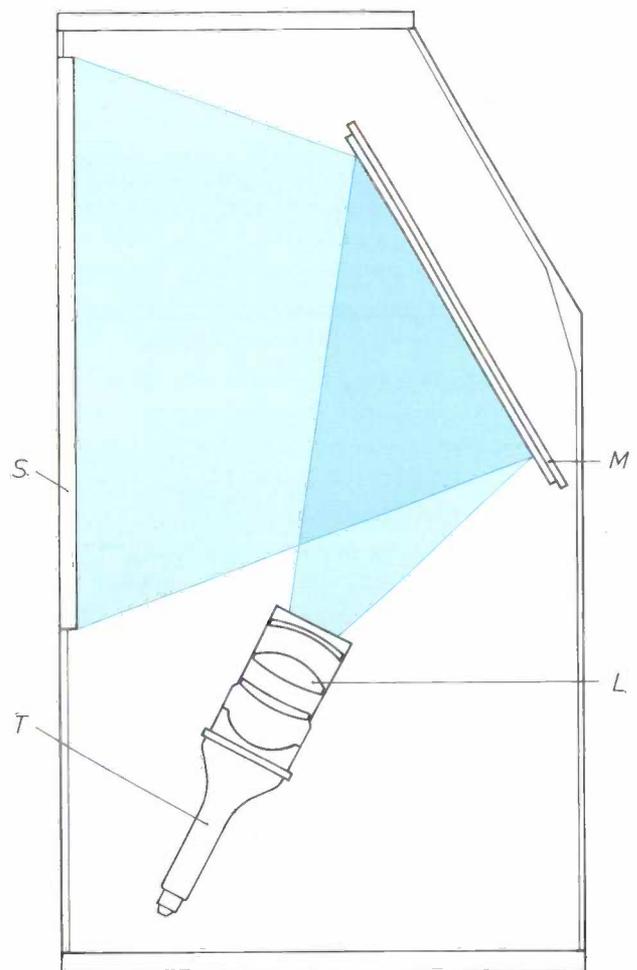


Fig. 1. Diagram of a rear-projection television set. The lens system *L* projects the picture on the faceplate of the tube *T* on to the viewing screen *S*. The mirror *M* provides a folded light path to keep the cabinet depth small.

tains three tubes, each with a different phosphor (to produce the red, green or blue image) applied to the faceplate. Projection of the three images in register on the screen gives the colour picture. Three lens systems are used, one for each tube. The components mentioned above, combined with the electronic control of the tubes and the mechanical and thermal stability of the mounting, determine the picture quality.

The picture on the viewing screen is magnified 8 to 10 times with respect to the picture on the faceplate. This means that the picture quality on the faceplates of the tubes must be very high; the brightness and sharpness are particularly important^[12]. The brightness (luminance) of the picture on the faceplate must be high to produce a brightness on the screen comparable with the brightness of a direct-view set. The brightness of the image on the screen decreases as the square of the magnification. The brightness also decreases for another reason; most of the light that leaves the faceplate falls outside the acceptance angle of the lens system and is lost; see *fig. 2*. The lenses used in projection television have a relative aperture of $F/1.0$ or less. When used at a magnification of 10 times a lens with such an aperture picks up light in a cone with a half-angle of 27° . This means that only 20% of the light originating from the tube passes through the lens to contribute to the brightness on the screen. With most lenses even less light is accepted from points away from the centre of the picture. There are also transmission losses (≈ 40 to 50%), mainly as a result of reflection and scattering at the surfaces of lenses, mirrors and projection screen.

As a result of the magnification, limited acceptance angle of the lens, and transmission losses, the brightness is reduced by a factor of 750 in projecting the image of one of the faceplates on to the screen. But, because the projection screen concentrates the light in a limited range of viewing angles, we recover a factor of five in brightness. Another factor of five is recovered by using three separate colour tubes instead of one shadowmask tube^[13]. This means that if the projection tubes are run with the same loading on the phosphor as in a direct-view tube, a direct-view television would be about 30 times brighter than a projection television. To bring the brightness of the projection television up to that of a direct-view television means that the brightness of the phosphor would have to be increased by a factor of 30. The 'green' phosphor used in direct-view colour television, for instance, cannot be used since it saturates before reaching this high brightness. Therefore other green phosphors have to be used, but these have a lower efficacy. For the green tube this has the result that the beam current must also be larger, by a factor of two. All these fac-

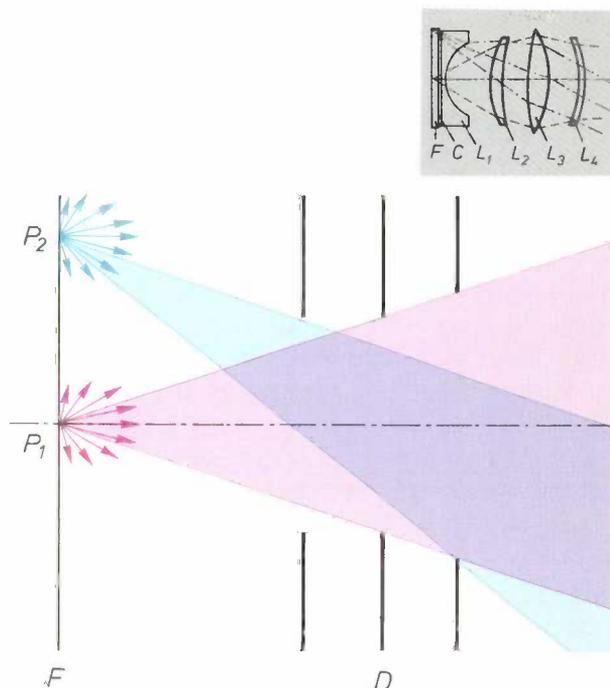


Fig. 2. Light acceptance by the lens system (schematic). When the electron beam strikes the phosphor layer light with a cosine intensity distribution is produced. This is shown for two points (P_1 and P_2) at the centre and the edge of the faceplate F of the tube by the lengths of the red and the blue arrows. Only some of this light (indicated by the red and blue shaded areas) is accepted by the lens system, represented schematically by the three diaphragms D . (More details of a typical lens system are shown in the inset diagram. It consists of 4 lens elements L_1 , L_2 , L_3 and L_4 . The first lens element is optically coupled to the faceplate by a liquid layer C .)

- [5] K. Ando, M. Osawa, M. Ohki, M. Ogino and E. Yamazaki, A 54-in. (5:3) high-contrast and brightness rear-projection display for high-definition TV, *SID Digest* **16**, 274-276, 1985; R. Bradley, Jr., J. F. Goldenberg and T. S. McKechnie, Ultra-wide viewing angle rear projection television screen, *IEEE Trans. CE-31*, 185-193, 1985.
- [6] R. L. Howe and B. H. Welham, Developments in plastic optics for projection television systems, *IEEE Trans. CE-26*, 44-53, 1980.
- [7] J. A. Clarke, Wide angle lenses for projection TV, *Proc. SPIE* **554**, 394-397, 1986.
- [8] R. Raue, A. T. Vink and T. Welker, Phosphor screens in cathode-ray tubes for projection television, to be published shortly in this journal.
- [9] T. G. Spanjer, A. A. van Gorkum, T. L. van Soest and M. R. T. Smits, A high-resolution electron gun for color projection CRTs, *SID Digest* **18**, 170-173, 1987. In one of the following issues of the Review attention will be paid to this research.
- [10] L. Vriens, J. H. M. Spruit, J. C. N. Rijpers, M. R. T. Smits and J. Khurgin, The interference filter projection TV CRT, *SID Digest* **19**, 214-217, 1988.
- [11] J. A. Clarke, J. H. M. Spruit and L. Vriens, Optical aspects of the interference filter projection CRT, *SID Digest* **19**, 218-221, 1988.
- [12] The sharpness of the picture on the faceplate is mainly determined by the characteristics of the electron beam. This will be discussed in the article of note [9].
- [13] In a direct-view tube, colour selection is usually obtained by using a shadowmask. The area of the openings is only 20%, and three electron beams pass through the mask (in slightly different directions, to land on the red, green and blue phosphor areas). So 60% of the faceplate area at most contributes to the luminescence. In projection television three single-colour tubes are used (red, green and blue), so we gain a factor of three divided by 0.6, i.e. a factor of five with respect to the brightness of a shadowmask tube.

tors together imply that the beam loading on the phosphor in projection tubes needs to be about 60 times larger than in a direct-view tube. A disadvantage of this high beam loading is that the phosphor layers on the faceplate degenerate much faster. Extensive research has therefore gone into new phosphors^[8] and techniques for applying them to the faceplates.

The reduction in brightness as described above applies to the centre of the picture, where the lens accepts 20% of the emitted light. As can be seen in fig. 2, much less of the light originating from off-axis parts of the faceplate of the tube is accepted by the lens and at the corners of the picture it may be as low as only about 5 to 7%. This happens because the limited size of the lens elements and parts of the mounting system, especially at the extremities of the lens, physically limit the off-axis acceptance cone to a greater extent than the on-axis cone. This effect is known as vignetting.

With all these problems, it is rather difficult to obtain a picture with sufficient brightness and resolution at the projection screen. To increase the amount of light accepted by the lens, it might seem that the acceptance angle of the lens ought to be increased. However, this is not a practical solution, because the aberrations^[14] of the lens increase too rapidly with increasing acceptance angle, resulting in a loss of picture sharpness. The lens would also become larger and more expensive.

In this article we describe a new solution to this problem; the use of an interference filter between the phosphor layer and the faceplate^[15], and the consequent need to make the faceplate curved. These two actions provide an elegant solution for the reduction of light losses and, rather surprisingly, improve the colour rendering, the resolution and the contrast at the same time.

An essential feature of the special type of multi-layer interference filters we used (short-wave-pass filters), is that the light emission from the faceplate is changed from an approximately Lambertian pattern (cosine distribution) in such a way that more light leaves the faceplate surface in the forward direction (at small angles α) and less light at large angles. More light therefore falls within the acceptance angle of the lens.

Another important feature of the interference filter is its wavelength selectivity, so that its use not only increases the brightness, but also removes light at unwanted wavelengths from the emitted spectrum. This improves the colour rendering and resolution in the projected image. We shall see in the next section how these improvements are obtained.

The multilayer interference filter

An interference filter consists of a number of thin layers of optically transparent materials with alternating high and low refractive indices^[16]. At each boundary between two layers there is transmission (with refraction) and reflection of the light waves. Reflection is associated with phase changes of 0 (light passing from material of high refractive index to material of low refractive index) and π (light passing from material of low refractive index to material of high refractive index). There are no phase changes on transmission. The optical (phase) thickness δ of a layer of geometrical thickness d is given by

$$\delta = \frac{2\pi n d \cos\alpha}{\lambda}, \quad (1)$$

where λ is the wavelength of the light, n is the refractive index of the layer and α is the angle between the direction of propagation of the light and the normal. If we choose the thickness of the layers so that

$$d = \frac{\lambda}{4n \cos\alpha}, \quad (2)$$

the optical thickness is $\delta = \pi/2$. Thus the phase difference between a light wave reflected at the first surface of the layer and a wave that is reflected at the second surface (which passes through the layer twice), is $2\delta = \pi$. When the phase changes at the interfaces of the high-refraction and low-refraction layers, 0 or π , are included as well, the maximum constructive interference occurs between the reflected waves. Similarly, with a directly transmitted wave and a wave reflected

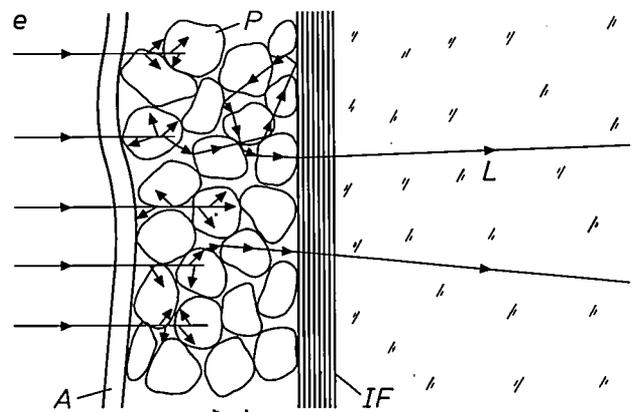


Fig. 3. Detail of the front end of the tube. The electron beam e excites the phosphor layer P and the emitted light quanta are transmitted if they are incident on the interference filter IF within an angle of about 36° to the normal. Light incident at a greater angle to the normal is reflected by the filter, is subsequently reflected by the aluminium backing A or scattered by the phosphor and can then enter the range of angles in which light L is transmitted.

at both sides of the layer, maximum destructive interference occurs on transmission.

By making use of a stack of such quarter-wave interference filters consisting of a sufficiently large number of layers — e.g. 20 — the reflectance for a single wavelength (λ) can be made high in one range of angles (α), with the transmittance high in another range of angles. We can therefore design an interference filter for use in a projection tube such that light at a wavelength λ equal to the central wavelength of the phosphor and incident within an angle of 30 to 40° to the normal is transmitted and light incident at larger angles is reflected. After scattering in the phosphor layer some of this reflected light comes within the range of angles where light is transmitted (see *fig. 3*). In this way the brightness of the light emitted within the acceptance angle of the lens increases, giving a gain in brightness of the projected image. At the same time, the amount of light outside the acceptance angle is reduced, resulting in less stray light. Stray light reduces the contrast in the projected picture.

Fig. 4 shows the transmittance of a short-wave-pass 20-layer $\text{TiO}_2/\text{SiO}_2$ filter as a function of the angle of incidence. The transmittance is high at small angles, while virtually no light is transmitted at large angles, thus concentrating the light around the normal to the surface. The exact shape of the transmittance func-

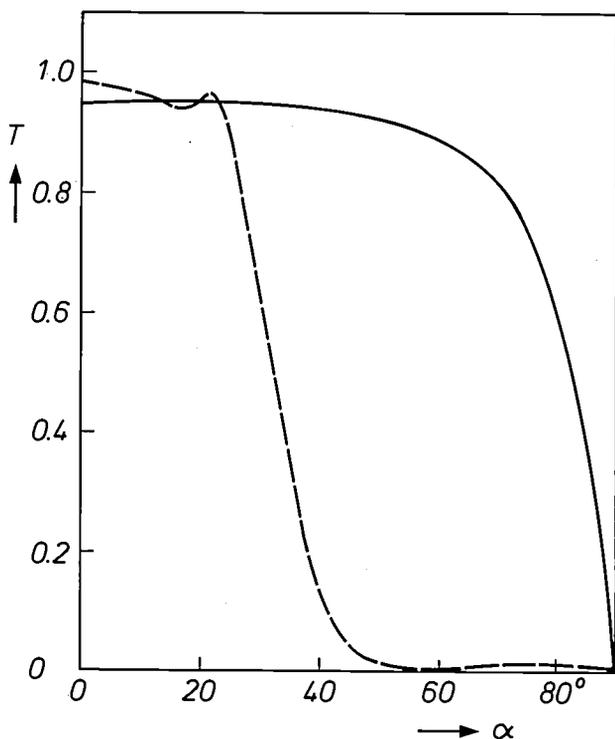


Fig. 4. Transmittance T of the faceplate with a 20-layer filter (dashed curve) and without a filter (continuous curve) as a function of the angle α (in vacuum) to the normal to the surface, for the central wavelength of the phosphor.

tion depends on the number of layers in the filter and on their thickness distribution.

We have just considered how the transmittance varies with angle (α) for a given thickness (d) and wavelength (λ). Equation (1) shows that the phase change (and therefore the transmittance) also varies with wavelength (λ). If the average thickness of the layers is matched to the central emission wavelength of the phosphor, light at the central wavelength is transmitted in the desired range of angles, while light at longer wavelengths is reduced in amplitude or may even be completely suppressed.

Fig. 5 shows the transmittance of an interference filter for a green-phosphor projection tube for perpendicular incidence of the light as a function of the

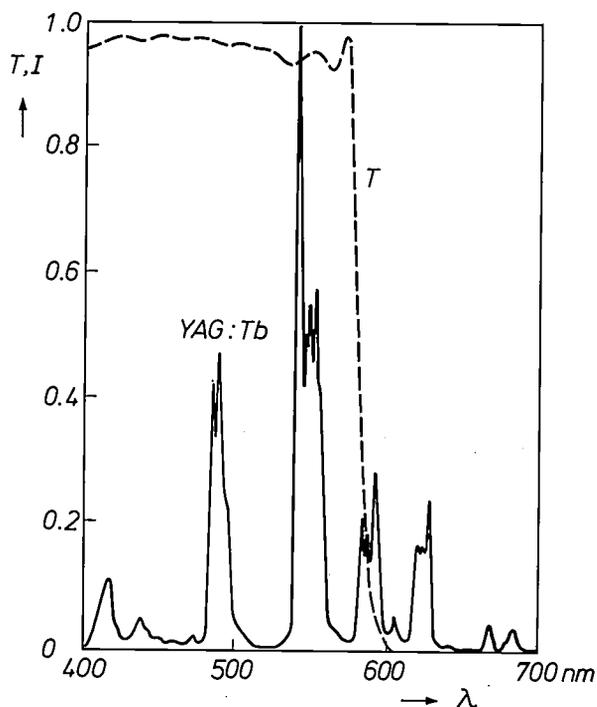


Fig. 5. Transmittance T of an interference filter as a function of the wavelength λ of the incident radiation. The spectrum of the green phosphor (I as a function of λ) is also shown in the figure. The filter shown will suppress the orange and red components in the spectrum of the green phosphor.

[14] The two most important aberrations are (a) spherical aberration, which causes a blurred image because the outer zones of the lens focus the light rays at a distance different from that of the central zone of the lens, and (b) chromatic aberration, which occurs because the refractive indices of the lens elements are dependent on the wavelength of the light. This gives different focal lengths at different wavelengths. The image of a point source of light with a broad spectral distribution is therefore blurred.

[15] The interference filter is applied to the inside of the faceplate, next to the phosphor layer, and not to the outside (which would be easier), because it is essential that light reflected by the filter is only scattered in the immediate vicinity of the originating point, otherwise serious degradation of the resolution will occur.

[16] More information on thin multilayer interference filters can be found in: H. A. Macleod, *Thin-film optical filters*, Adam Hilger, 2nd edition, Bristol 1986.

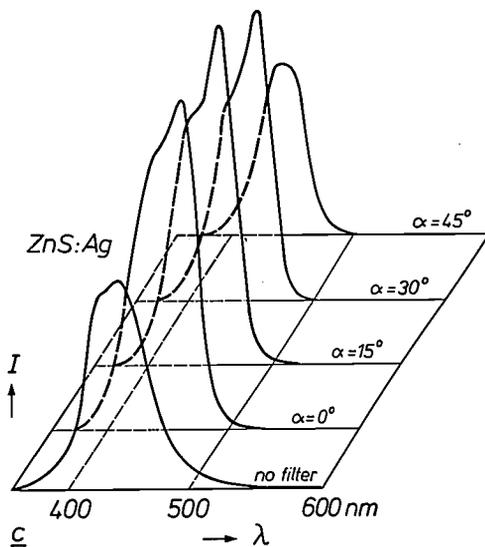
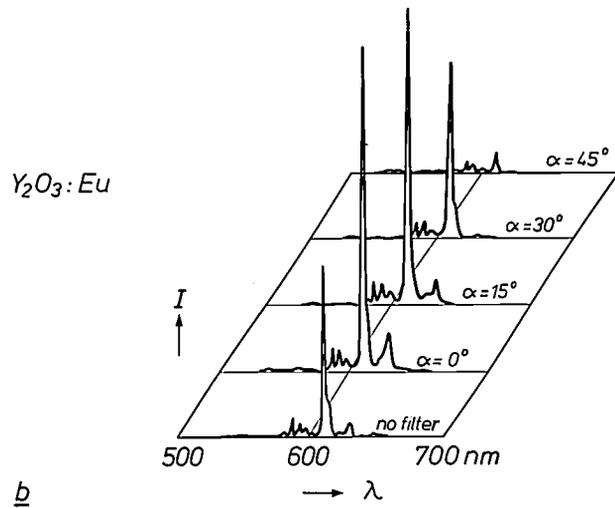
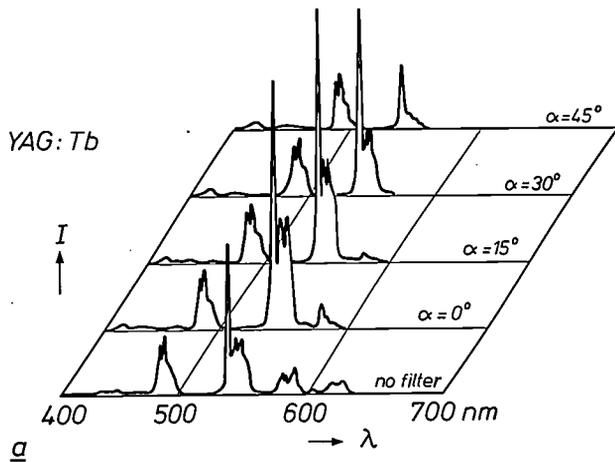


Fig. 6. The spectra (I as a function of λ) emitted by the green (a), red (b) and blue (c) tubes with filters, measured at various angles to the normal to the surface. The emission of tubes without a filter (almost independent of angle) is also shown.

wavelength. This figure also shows the spectrum of the green phosphor: $Y_3Al_5O_{12}:Tb$ (YAG:Tb). Besides the dominant emission centred on a wavelength of 545 nm, this phosphor emits light at other wavelengths, in the blue, orange and red. A problem with practically all green Tb-activated phosphors for projection television is that the orange/red contribution to the spectrum is too intense, so that green appears as a yellowish green rather than a true green. We have already seen that when a short-wave-pass interference filter is incorporated between the phosphor layer and the faceplate, the spectrum of the light beam emitted from the faceplate will also be dependent on the angle from the normal (see fig. 6a). At small angles the green component in the spectrum is increased, the blue is transmitted unchanged, the orange component is largely suppressed and the red component is not transmitted at all^[17]. The undesired contributions from the longer red and orange wavelengths to the green image are thus smaller and this improves the chromaticity.

For the red tube, which contains $Y_2O_3:Eu$, the spectral distribution (fig. 6b) shows a large gain at small angles and a loss at large angles. For the blue tube, which contains $ZnS:Ag$, fig. 6c shows that with no filter there is a broadband emission extending into the green, yellow and red. This long-wavelength tail gives rise to a significant chromatic aberration. With a filter there is a gain in the blue spectral region and the long wavelengths are suppressed. The chromaticity of the blue tube is therefore also improved by the application of an interference filter.

Taking the ratio of the zero-angle spectral distributions with and without a filter gives the zero-angle gain curves. In fig. 7 curves of gain as a function of the wavelength are shown for green, red and blue tubes with filters. Since the measurements of the spectral distributions with and without a filter were made in the same tube (the filter only covered half of the

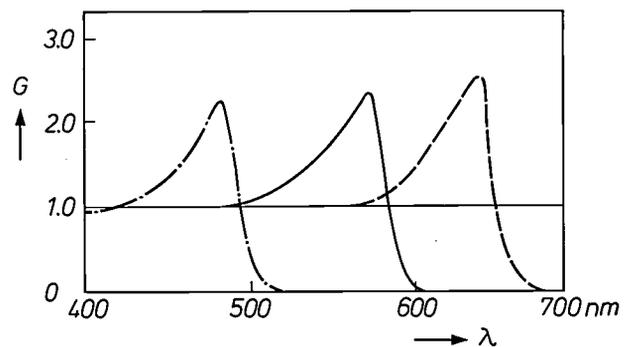


Fig. 7. The gain G as a function of the wavelength λ for the filters applied to the green (continuous line), red (dashed line) and the blue (chain-dotted line) tubes.

faceplate) these comparisons are independent of any differences between different tubes that might arise, for example because of different processing conditions. The maximum gain in the forward direction is 2.4 to 2.7. We shall see below that such large gain factors cannot be obtained in projection sets for the home, but that a factor of 1.5 can be obtained.

Quality of the projected picture

The principles and many of the features of multilayer filters as used in projection tubes have been discussed in the previous section. We shall now look more closely at the effects of these filters on the projected picture, and at the modifications that have to be made to other components if they are to work in a practical projection set. First we shall consider the brightness distribution on the viewing screen. Other aspects of picture quality affected by the use of a filter — chromaticity, resolution and contrast — are discussed in the last section.

Brightness on the viewing screen

We have seen that a short-wave-pass multilayer interference filter on the inside of a tube faceplate alters the angular distribution of the emitted light so that more light is emitted at small angles and less at large angles; see *fig. 8a*. *Fig. 8b* shows the measured brightness on the viewing screen when 5-inch diagonal pictures, with and without such a filter, are projected at a magnification of 8 times with a lens of relative aperture $F/1.2$ and focal length of 135 mm. Without a filter the brightness decreases as a function of the distance to the centre. With a filter there is still a difference in brightness between the central and the non-central parts of the picture, but there is a gain in brightness for the entire projection screen.

Problems arise, however, when filters are used in projection sets for home use. Lenses with focal lengths of 80 to 100 mm are normally used here, as the focal lengths must be small to keep the cabinet size small. For points in the picture near the axis the filter does indeed have the desired effect: more light is emitted in the forward direction. For points in the picture away from the axis the situation is different. As can be seen from *fig. 8a*, for points well away from the axis the lens only picks up light at large angles, so only a small fraction of the light in the new distribution, which is centred on the normal to the faceplate, comes within the acceptance angle of these lenses. This means that using the filter increases the brightness at the centre of the picture, whereas the brightness at the edges of the picture decreases (see *fig. 9*). It can be seen that the greater the on-axis gain, i.e. the smaller

the cut-off angle $\alpha_{0.5}$, the greater the loss at the edges. This is unacceptable, of course.

There are two approaches that can overcome this problem. The first is to design a lens in which the off-axis acceptance cone does not tilt: such lenses are more complex, more bulky, and more expensive. The second approach is to modify the tube so that the nor-

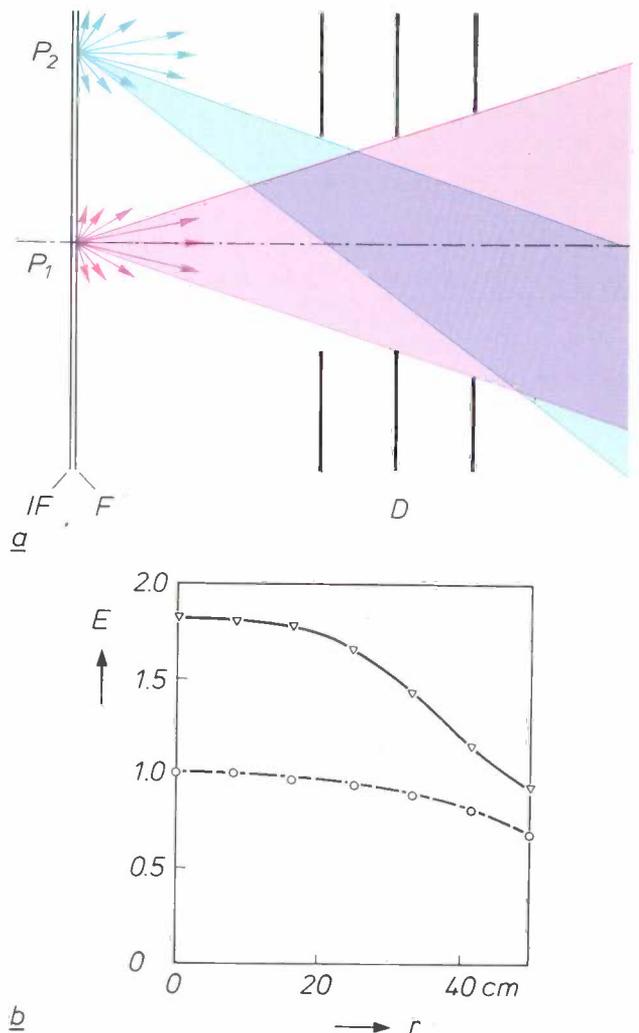


Fig. 8. a) When the faceplate F of the tube is given a multilayer interference filter IF more light is emitted in a forward direction. This is indicated for two points P_1 and P_2 at the centre and at the edge of the faceplate by red and blue arrows. The lengths of the arrows are an indication of the light-intensity distribution. Light accepted by the lens system, represented schematically by three diaphragms D , is indicated by the red and blue shaded areas. b) Brightness E measured on a projection screen when a phosphor layer emitting a uniform brightness is projected from a tube with (continuous line) and without (dashed line) an interference filter as a function of the distance r from the centre of the screen. The tube had a flat faceplate and the red phosphor $Y_2O_3:Eu$. The lens used for projection was a $F/1.2$ lens with a focal length of 135 mm. The brightness values are normalized to the brightness at the centre without a filter.

[17] The interference filter is matched to the central emission wavelength of the phosphor to give a cone of transmittance angles from 0° to 30° or 40° . At the longer wavelengths this cone becomes more narrow until light is no longer transmitted at small angles and at shorter wavelengths it widens to give less reflection of light and hence less gain.

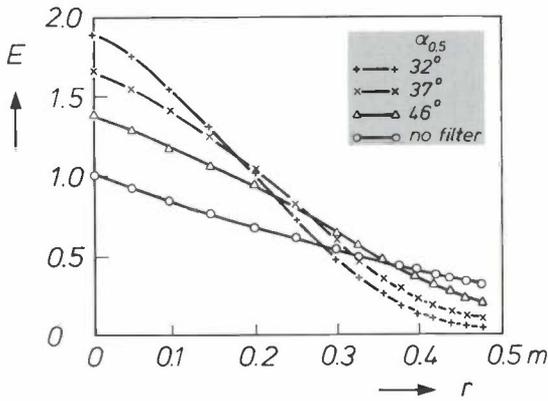


Fig. 9. Relative brightness E as a function of the distance r from the centre of the screen of a projection television, when an image of a flat faceplate of uniform brightness is produced. The focal length of the lens is 90 mm. Interference filters with different cut-off angles $\alpha_{0.5}$ (defined as the angles at which the filter transmittance is 50% for a particular wavelength) are applied to the faceplate. The curves are normalized to the brightness at the centre without a filter.

mal to the faceplate is directed towards the lens acceptance cone for on-axis and off-axis points. This can be done by using a curved faceplate instead of a flat one. The effect can be seen in *fig. 10a*, and results in increased brightness throughout the picture (see *fig. 10b*).

For a complete match between tube emission angles and lens acceptance angles, the radius of curvature of the faceplate would depend on the type of lens, and especially on its focal length. In projection sets for the home, where lenses with fairly short focal lengths (≤ 90 mm) are used, the ideal radius would be 200 mm or less. It is not easy to deposit the layers of the filter and the phosphor on such steeply curved faceplates, and dynamic focusing of the electron beam in the tube would also become too much of a problem. It is however possible to arrive at a compromise, and for tubes with a 5-inch diagonal picture the filter could be applied to a faceplate with a radius of curvature of 350 mm. A filter can then be chosen such that the brightness is about 1.5 times higher at the centre of the projected image, while the situation at the edges is much better than when a flat faceplate is used; see *fig. 11*. In High Definition Television (HDTV) sets, lenses of longer focal length are often used. The advantages of using a multilayer filter on a curved faceplate are greater, and gains of 2 are feasible.

Chromaticity, resolution and contrast

In addition to making better use of the interference filter in off-axis parts of the picture, the introduction of curved faceplates has other consequences for the optical system. An advantage is that the first element of the lens does not have to be so powerful. This

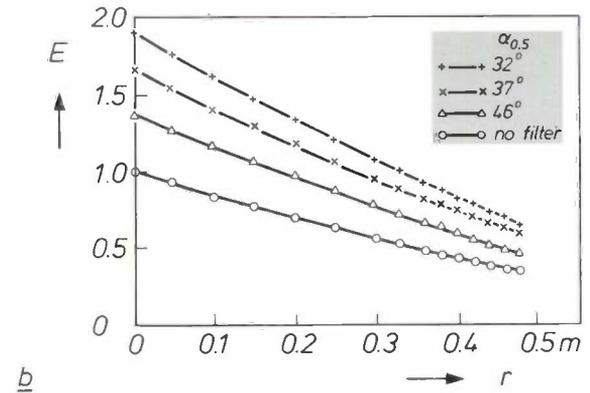
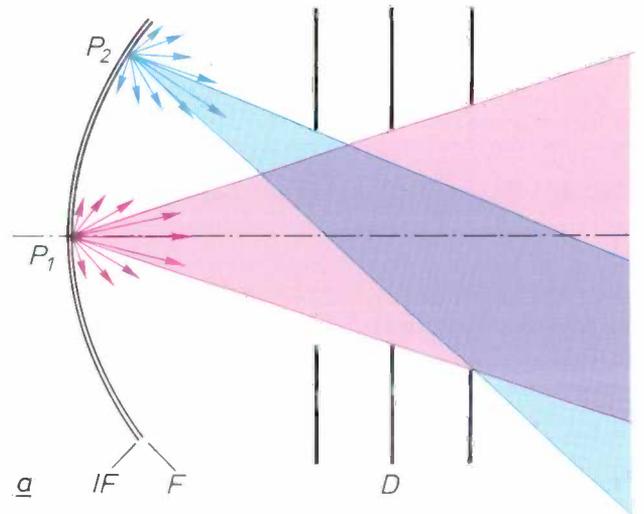


Fig. 10. *a)* A multilayer interference filter IF applied to a strongly curved faceplate F changes the light distribution (for two points P_1 and P_2 indicated by red and blue arrows (see *figs 2 and 8a*)) in such a way that much more light is accepted from the edge by the lens system (represented schematically by the diaphragms D). *b)* Relative brightness E as a function of r for different cut-off angles $\alpha_{0.5}$ (as in *fig. 9*) when a curved faceplate of radius of curvature 200 mm and with a uniform brightness is projected on the screen.

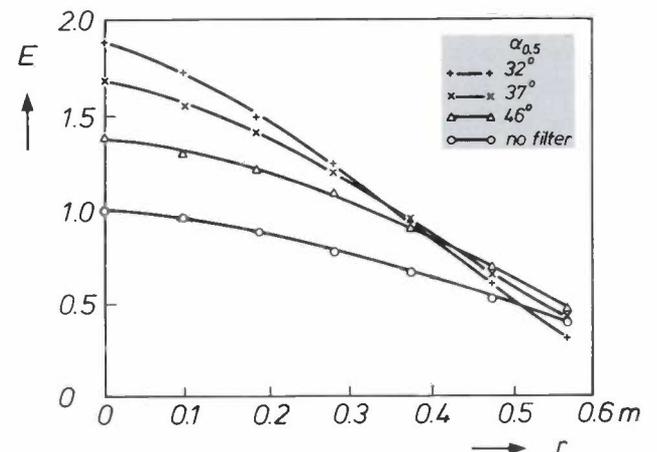


Fig. 11. Relative brightness E as a function of r for different cut-off angles $\alpha_{0.5}$ (as in *fig. 9*) when a curved faceplate of radius of curvature 350 mm and a uniform brightness is projected on the screen.

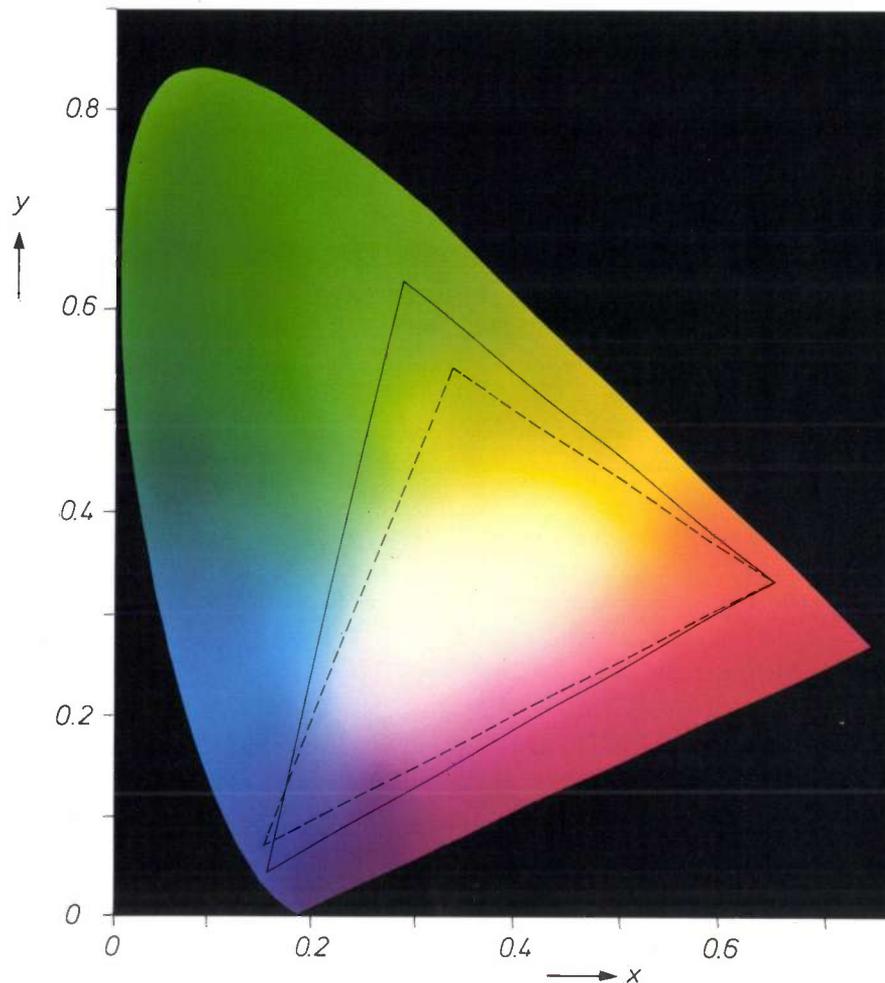


Fig. 12. C.I.E. chromaticity diagram in which the chromaticity coordinates x and y are the horizontal and vertical axes. The area inside the dashed triangle shows the colours previously available in projection sets. The corners of the triangle give the chromaticities of the red, green and blue phosphors. The continuous triangle shows the colours available when interference filters are used. The triangle is larger with filters, giving a wider colour gamut.

means that the off-axis rays do not meet the surface at such severe angles of incidence, with a consequent reduction in the aberrations in the lens, especially the astigmatism. This in turn increases the resolution of the lens and reduces the need for vignetting, so further increasing the brightness of the projected picture. Improved optical performance is in fact obtained when a curved faceplate is used without an interference filter, as had been noted in the early days of projection television^[2]. A disadvantage is that the curved faceplate introduces optical distortion into the picture^[18], and this has to be corrected electronically in the drive circuits for the tube.

The chromaticity is also important in evaluating the picture quality of a projection television. Colour is directly related to the wavelengths of the light emitted. The perception and evaluation of colour is very sub-

jective. An objective standard for colour quality is defined by the CIE 'chromaticity coordinates' x and y , which are related to the perception of colour by large numbers of observers^[19]. (White is indicated by $x \approx \frac{1}{3}$, $y \approx \frac{1}{3}$.) In the x, y -plane, areas can be designated where colours meet certain specifications. In *fig. 12*, the colour coordinates for tubes with and without interference filters are indicated. The interference filter has an effect on the spectral distribution of the light leaving the faceplate, as shown earlier (*fig. 6*). In the spectrum of the green phosphor the filter suppresses the red and orange components of the light and x becomes smaller and y larger. The dynamic

^[18] J. A. Clarke, Current trends in optics for projection TV, *Opt. Eng.* 27, 16-22, 1988.

^[19] See for example R. W. G. Hunt, *Measuring colour*, John Wiley, Chichester 1987.

colour range is thus greater when filters are used. The chromaticity coordinates for the light emitted from the green tube are brought closer to the European Broadcasting Union^[20] recommendations. Similarly, in the spectrum of the blue phosphor, the long-wavelength tail into the green and the red is suppressed completely and there is a gain in the blue part of the spectrum. The spectrum of the red phosphor is almost monochromatic and close to or within the EBU specification, so it is only necessary to design the filter to have its highest gain at the central wavelength. The improved chromaticity of the green signal makes a significant contribution to improved colour rendering in the picture.

A direct consequence of the narrower spectral distributions of the light leaving the faceplate is that the effect of chromatic aberration^[14] is also reduced. Light with a broad spectral distribution and originating from a single point of the faceplate will be projected on the screen as an enlarged spot with the colour changing from centre to edge. When a filter is used, the width of the spectral distribution is reduced, giving a smaller spot and hence better resolution. This reduction in spot size is far more important than the slight increase in spot size on the faceplate due to reflection of light between the filter and phosphor layer: the light responsible for the gain in brightness.

Another consequence of less light leaving the faceplate at large angles is that less light reaches the edges of the lens elements and the lens housing. Ideally all of the light that misses the lens pupil should be absorbed by the mechanical components that support the lens elements and the tube. In practice, a significant amount is scattered back to the phosphor but away from its point of origin, reducing the contrast in the projected picture.

We have shown that multilayer interference filters used on curved faceplates in tubes for projection television increase the brightness at the screen. As well as the brightness, other aspects of the picture (chromaticity, resolution and contrast) have also been improved. The improvements in projection tubes (phosphors and electron guns), lenses and viewing screens have already contributed to the greater market share for projection television. The introduction of projection sets with curved-faceplate tubes and interference filters on the American market in early 1989 should give a further increase in the market share. This article describes the first two stages in the creation of this new product.

The authors wish to thank J. C. N. Rijpers, A. M. J. Roosen and H. M. de Vrieze for making the interference filters, M. R. T. Smits for performing the recent tube experiments, J. Khurgin for his early work at PRLN, Briarcliff Manor, R. Raue and T. Welker and their colleagues for applying the phosphor layers to the tubes, and the workshop at Philips Research Laboratories in Eindhoven for manufacturing the tubes.

^[20] E.B.U. standard for chromaticity tolerances for studio monitors, Tech. 3213-E, European Broadcasting Union, Brussels 1975.

Summary. A multilayer interference filter — a stack of layers of alternately high and low refractive indices — alters the distribution of the light emitted by the phosphor on the faceplate of a television tube in such a way that more light is emitted in the forward direction. When these filters are used in projection television tubes, more light falls within the acceptance angle of the lens system, so that more light contributes to the brightness on the projection screen. The brightness everywhere on the screen can be improved by using curved faceplates in the tubes. The chromaticity, resolution and contrast in the projected picture are also improved when such filters are used.

1939

THEN AND NOW

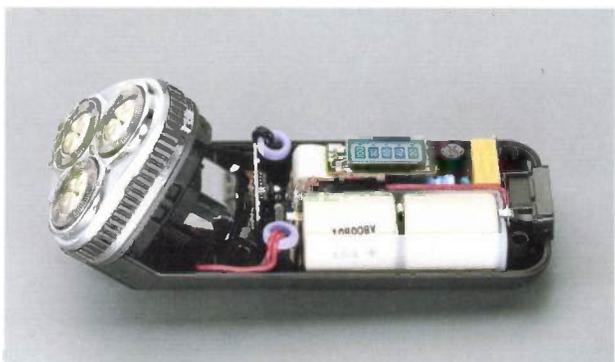
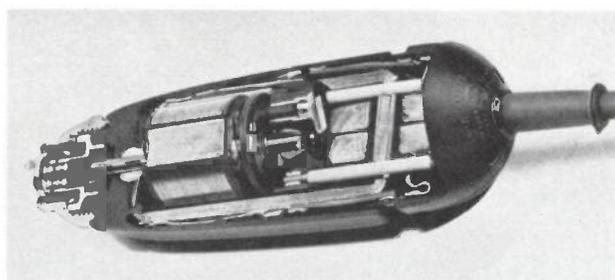
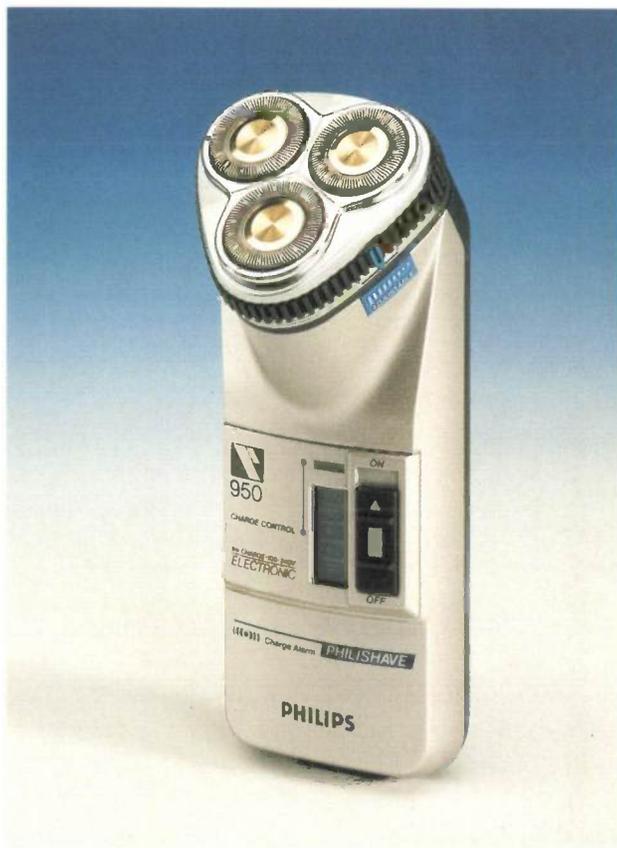
1989

Philishave

Fifty years ago Alexander Horowitz, the 'Father of the Philishave', described how it all started: a shaver head with slots in all directions captured the bristles, which were then shaved off by rapidly rotating cutters. The first Philishave, the 'Staalbaard' (it means 'steel beard' in Dutch; see black-and-white photograph [*]), had three bronze cutters. The rotary movement of the cutters ensured a constant high speed, unlike other systems with vibrating cutters.

The principle of the Philishave has remained unchanged through the years. But refinements in the manufacturing process and a better understanding of the behaviour of bristles and skin have brought continuous improvement. The number of slots in the head has increased from 48 to 90, for example, and today there are three heads instead of one. The heads are now spring-mounted, with height adjustment to suit different skins. In 1980 there was a particularly notable improvement in the quality of the shave with the introduction of a thin and light second cutter blade in front of each main cutter blade. This 'lift blade' raises the bristle slightly, so that the main blade can shave it off even more closely.

The Philishave owes much of its considerable success to the close attention to user comfort and its



attractive design — see the colour photograph of the HP 950, the latest model. It is lighter, smaller and quieter; it is easier to hold, has rechargeable batteries and adapts automatically to the mains voltage.

The large series motor (black-and-white photograph) has been replaced by a much smaller and lighter motor with permanent magnets (second colour photograph). This motor gives the designer much more freedom for styling. An electronic circuit controls power supply to the motor and battery charging, and shows the state of charge on an LCD. It also notes the individual energy consumption: the unused shaving time is shown as a percentage of the time available with a full charge.

Horowitz would never have guessed in 1939 that his Philishave — known as the Norelco Shaver in the United States — would now be the world's most successful shaver.

[*] From Philips Technical Review, December 1939.

Analysis of the injection-moulding process

J. F. Dijksman

To mark Dr E. A. Muijderman's retirement a colloquium was held at Philips Research Laboratories, Eindhoven, on 27th May 1988. All the subjects discussed were closely associated with Dr Muijderman's work in science and engineering — work that has included research on plastic products and has provided valuable spin-off to other research. This was evident from one of the talks at the colloquium, in which Dr Dijksman examined the injection-moulding process used for the manufacture of plastic products. The text given below is closely based on the talk.

In very many cases we can say that the interface between the user of a Philips device and its function is a plastic product. The casing of the Philishave, for example, is made of plastic, sometimes finished with a metal-coloured coating to make it look like metal. The outer casings of coffee makers, kitchen machines, irons, vacuum cleaners, and sometimes parts inside, are formed from plastic components. Very complex shapes can be made (*fig. 1*), and this means that a number of functions can be integrated, such as

- load distribution,
- shielding rotating parts and electrical connections,
- sound insulation,
- replacement of fasteners such as screws by snap connectors,
- design.

One component that has to take a very heavy load is the plastic tub of a washing machine; see *fig. 2*. Another highly stressed component is found in the Philips top-loader washing machines. This is the internal top frame of fibreglass-reinforced plastic from which the drum is suspended. The control knobs of amplifiers, tuners, Compact Disc players, record players and so on are made of plastic. In a remote-control unit for a

television set the integration of functions has progressed so far (pushbuttons, guide and return-spring mechanism for the pushbuttons, snap connectors between top and base) that it only consists of a few plastic components, LEDs (light-emitting diodes) and a number of electronic components. This is a variation on the integrated-circuit theme, for here we have mechanical components in which various functions have been integrated.

To generalize again, whenever you encounter plastic components in our products, you will usually find that these components have been made by injection

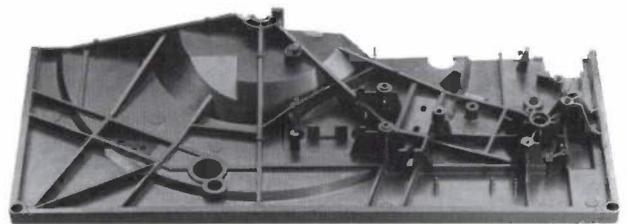


Fig. 1. The turntable of a record player is an example of a plastic product with a complicated shape. The photograph shows a half of a turntable, seen from below.

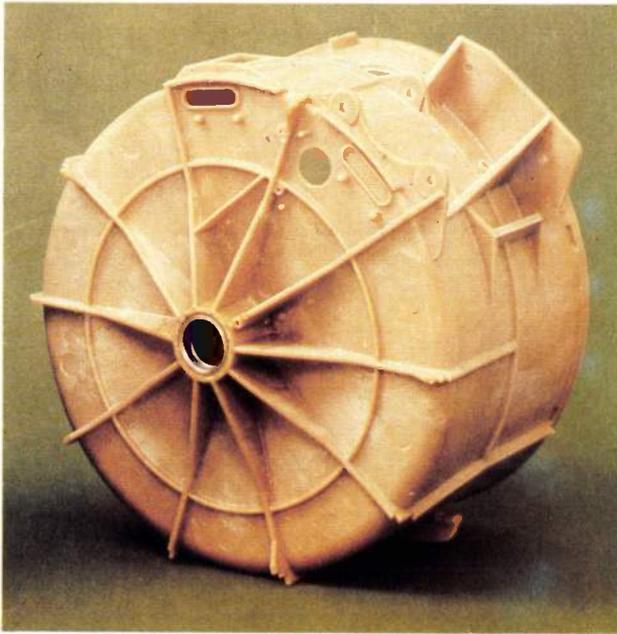


Fig. 2. The tub of a washing machine has to take a very heavy load. It is suspended in a frame and holds the water used in the wash. The bearings for the drum go through one side of the tub. It also has blocks of iron or concrete to increase the mass so that the machine does not move during the spin.

ly from electrical cuff heaters wrapped round the cylinder. While it rotates the screw moves slowly backwards (to the right), so that the molten material collects in the space formed in front of the screw. As soon as sufficient molten material has collected, the screw stops rotating. The screw, which now takes on the role of a plunger, is next pushed forward, forcing the molten material into the mould. Flowback is prevented by a non-return valve or the resistance along the screw. When the mould is full, the material has to cool until the product has become strong enough to be removed. Meanwhile new material for the next product is being melted and transported.

I should now like to explain the injection-moulding process from a number of different viewpoints — as it is seen by

- the materials scientist,
- the designer of the component or product,
- the mould designer,
- the rheologist.

So far I have been talking about plastics in general. The materials scientist knows however that many different types of materials lurk behind that word 'plastics', and I shall name some of them.

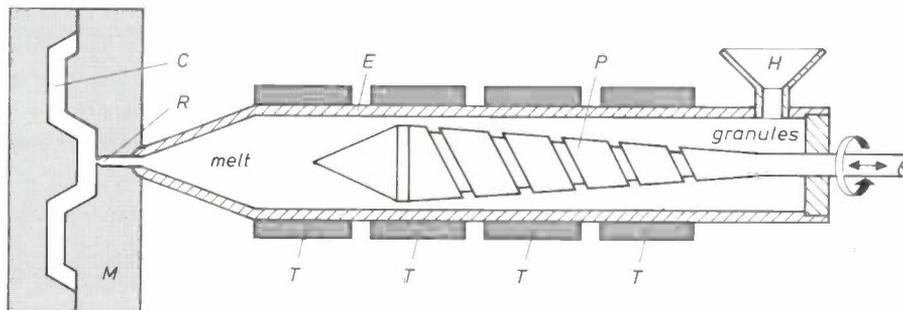


Fig. 3. Injection-moulding process, schematic. Plastic granules fed via the hopper *H* to the cylinder *E* are transported towards the runner *R* by the plunger screw *P* as it moves to the right. The granules are heated during this transport by internal friction and the electrical cuff heaters *T*. When sufficient melt has collected to its left, the screw is used as a plunger to inject the plastic into the cavity *C* of the mould *M* in a single stroke.

tion moulding. This process is widely used for the mass-production of plastic devices and components. (The same process is used for metal, too, but then it is called pressure die-casting.) The shaping of the component takes place during this process almost always in the following way. The original material, granules of the plastic to be used, is melted in the injection unit, the 'extruder', which is a cylinder containing a plunger screw; see fig. 3. The heat required is supplied partly from the friction between the granules as they move as a result of the rotation of the screw and part-

Amorphous plastics. Amorphous materials have the same structure in the liquid phase and the solid phase (there is no crystalline structure) and are often transparent. Well-known transparent plastics in this group are polycarbonate (PC), which is used as a substrate for the Compact Disc, and polymethyl methacrylate (PMMA), which is used as the substrate in the LaserVision disc; see fig. 4.

Partially crystalline plastics. These materials have a fairly complete crystal structure in the solid phase. As in metals the crystal structure is highly dependent on

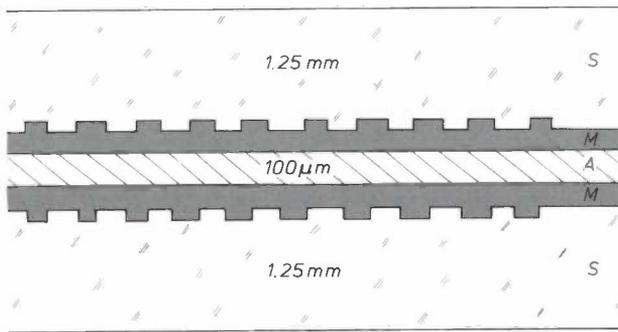


Fig. 4. Schematic configuration of a double LaserVision disc. *S* transparent substrate of polymethyl methacrylate (PMMA) that carries the picture and sound information in the form of pits. *M* metal reflecting coating, 0.6 μm thick. *A* adhesive layer, 100 μm thick.

the previous history of temperature changes. Polypropylene, one of the few coffee-resistant plastics, belongs to this group, as do polyethylene and polyvinyl chloride, the plastics that are widely used for the insulation of electrical wire.

Monomers. The plastics named above are polymer materials. In a number of cases, however, we start from monomers and carry out the polymerization in the mould under the influence of temperature or light. The vulcanization of rubber is an example of such a process. Materials such as phenolformaldehyde (Philit), some kinds of nylon, polyurethanes and most polyesters are also examples of materials that polymerize during the shaping process.

Mixtures (Blends). To improve the properties of plastics or adapt them to specific requirements many plastics are supplied as mixtures of different materials. There are many examples of this. Rubber particles are dispersed in polystyrene to increase its impact strength. The addition of large quantities of sand reduces the coefficient of expansion. It is used in encapsulations for ICs and in precision products like the 'arms' in Compact Disc players and for bearings. Another example is the addition of fibreglass to increase stiffness and strength. Finally, I should mention the addition of talc: this is done to lower the price.

The designer of the product or component has laid down a number of specifications for it. These might include

- mechanical strength and stiffness (how much load can the component take and how much is it allowed to 'give' under this load),
- creep (how much may the material flow under continuous load),
- temperature range,
- sensitivity to moisture, electrical insulation, dielectric properties, magnetic properties,

- colour,
- resistance to chemicals and radiation,
- shape stability (how large are the tolerances),
- toxicity, flammability,
- re-use (can it be recycled).

Working from these specifications, the designer endeavours to choose shape and material so that the product has the required properties. In practice this means that he consults the materials scientist.

In choosing the original material account must be taken of the fact that the material does not only have intrinsic properties. A number of properties can change because of the shaping process. This also means, of course, that these properties can be affected. They include

- optical properties (in Compact Discs, for example, the birefringence is affected by molecular orientation and inhomogeneous cooling),
- mechanical properties (as a result of the molecular orientation properties such as stiffness, strength and impact resistance may be anisotropic; inhomogeneous cooling can lead to internal stresses, resulting in reduced load-bearing strength),
- thermal properties (the thermal conductivity and the coefficient of expansion may become anisotropic),
- density (this may vary inside the product or component),
- the degree of crystallization,
- microstructure.

This list of properties that can be affected to a greater or lesser degree clearly shows the need for research into the relationship between shaping and product properties.

This research on plastics, which was mostly done on the actual components or products, provided valuable spin-off to other research at Philips Research Laboratories, e.g. on plasticity of materials^[1], flow mechanics in molten polymers^[2] and the mechanical aspects of polymer materials, particularly in relation to precision injection moulding^[3].

We now consider the matter from another viewpoint, as the mould designer sees it. The mould is the unit containing the cavity, and is mounted on the cylinder that contains the reciprocating screw. The choice of product geometry and material lands on the desk of the mould designer, who finds he has to answer the following questions:

- How should I choose the dimensions of the cavity so that I can compensate for the thermal shrinkage of the component? How can I prevent the component or product from warping and how can I avoid unwanted anisotropy?
- What will be the consequences of forced cooling? In general the product will not just be left to cool down,

since the economics requires the shortest possible cycle time; therefore the product must be cooled as quickly as possible. Certainly in the vicinity of the walls there is more likely to be 'quenching' rather than unaided cooling, with all the associated consequences for the structure of the injected material; extra shrinkage, warping and anisotropy can arise.

- How should I choose the stiffness and strength of the mould so that it does not distort too much under the influence of the forces resulting from the high pressure in the cavity. Sometimes the distortion can be put to use to compensate for shrinkage.

- Where should I put the gates? A gate is the connection between the runner channels in the mould and the cavity. When you buy a model aeroplane, you get the 'runners' as well. The gates are where you break off the parts of the aeroplane. Together with the injection temperature — the temperature of the melt as prepared by the extruder — and the injection rate, the locations of the gates determine the flow pattern in the cavity, and therefore the fill time and the pressure required. Provided the material in the gate has not solidified ('frozen'), material can be added during the cooling to compensate for shrinkage (packing). Otherwise, the positions of the gates are partly determined by aesthetic considerations (e.g. the position of a weld line, i.e. the line where the flows of material from different gates first meet). The flow pattern determines the anisotropy in the outer layer of the product.

- How should I design the layout of the runner system (*fig. 5*)? The idea is to have as little waste as possible, with balanced injection and packing.

- How should the mould be divided so that it can be opened and the product ejected (*fig. 6*)?

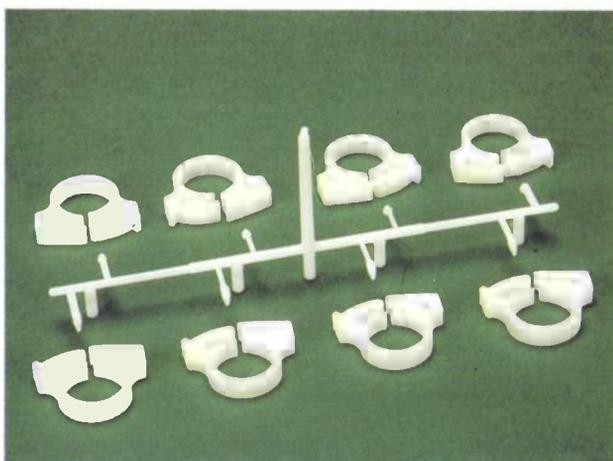


Fig. 5. Eight spring clips with their 'stalks' or runner cores, which reproduce the geometry of the runner system. The runner system consists of the long vertical feed and the horizontal channel with eight branches. (The eight vertical branches are used in ejecting the product from the mould.)

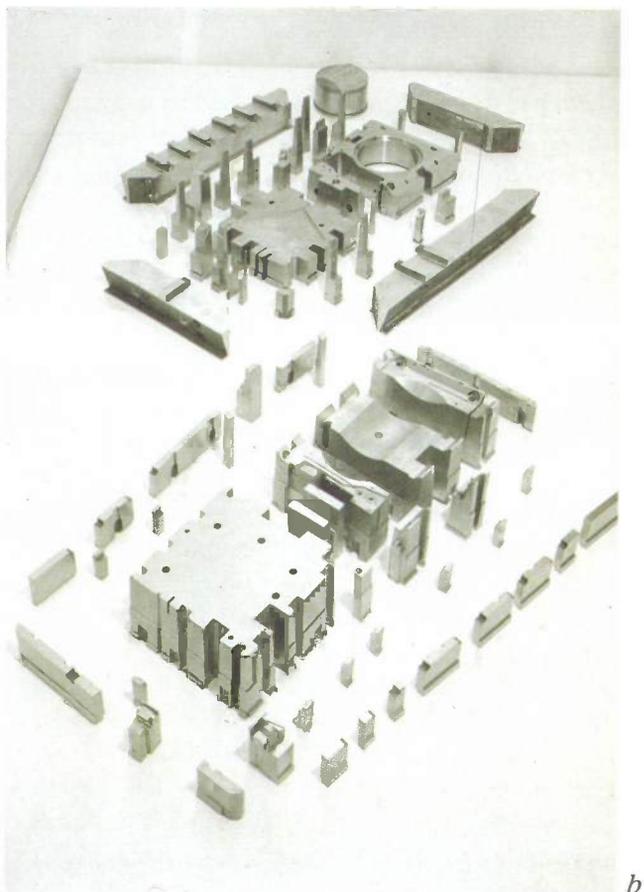
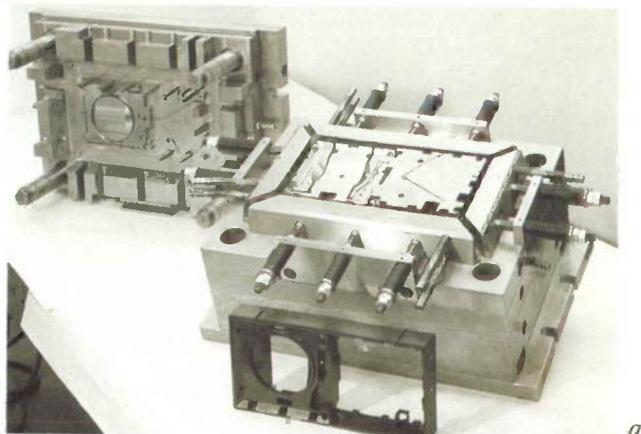


Fig. 6. *a)* Splitting the mould to free the product — part of the radio cabinet in the foreground. *b)* In practice each of the two parts of the mould is built up from a large number of individual components. In the photograph the lower part of the mould has been dismantled.

- [1] H. van Wijngaarden, Constitutive equations for metals with an application to the extrusion of lead, Thesis, Eindhoven 1988.
- [2] A. A. M. Flaman and B. Veltman, Injection moulding experiments, a challenge to numerical simulation programs, suppl. to Rheol. Acta 26 (Proc. 2nd Conf. of Eur. Rheologists, Progress and Trends in Rheology II, Prague 1986), 129-131, 1988.
- [3] F. P. T. Baaijens, Compressible solidifying flow of a molten polymer, in: A. W. Bush, B. A. Lewis and M. D. Warren (eds), Flow modelling in industrial processes, Ellis Horwood, Chichester, to appear shortly.

Finally, let us have a look from the viewpoint of the rheologist. Rheology is the science of the deformation and flow of matter due to external forces. So what does the rheologist see? To start with, he sees a material with a complicated behaviour. Complicated, because

- the coefficient of expansion is high,
- the mechanical properties depend closely on the temperature and sometimes on the time in service,
- the viscosity in the molten state is extremely high; this must be properly taken into account during the shaping (water at 20 °C has a viscosity of 10^{-3} Pa s, for example, while polystyrene at 250 °C has a viscosity of about 10^4 Pa s),
- the material is viscoelastic in both the solid phase and the liquid phase,
- most engineering plastics are extremely poor conductors of heat (diamond has a thermal conductivity of $2000 \text{ Wm}^{-1}\text{K}^{-1}$, copper $300\text{-}400 \text{ Wm}^{-1}\text{K}^{-1}$, steel $40 \text{ Wm}^{-1}\text{K}^{-1}$ and polystyrene $0.13 \text{ Wm}^{-1}\text{K}^{-1}$).

This last property must also be properly taken into account in the design of plastic products. In practice, because of the cooling time, only thin-walled products (fig. 7) or very small products can be manufactured economically from plastics.

We saw earlier that the product designer has to cooperate closely with the materials scientist, and here — because of the rheological behaviour of the material in the mould — it turns out that the rheologist and the mould designer will have a great deal to discuss. In fact, of course, a good product will only result from a cooperative effort by all concerned: the materials scientist, the product designer, the mould designer and the rheologist.

The title of my talk is 'Analysis of the injection-moulding process'. So far I have drawn your attention to various aspects of the material, the product design, the mould design and the rheology. A good product or component not only meets its specification, but can also be manufactured at an acceptable price. This means that the design of that component or product and the choice of the manufacturing process must go hand in hand. Computer aids can be very useful here.

The use of a flow-modelling program^[4] such as INJECT-3^[5], developed at the Philips Centre for Manufacturing Technology, will give the product designer, the mould designer and the rheologist a better understanding of the flow pattern in the cavity while it is being filled. This in turn will provide a clearer picture of the expected orientation of the molecules and any anisotropy in the product (prediction of the fill pattern). A knowledge of various quantities is required here, such as viscosity/viscoelasticity, thermal conductivity, specific heat capacity, compressibility,

coefficient of expansion. As yet the program can only be used for amorphous thermoplastics. In partially crystalline materials matters such as latent heat, crystallization kinetics and the effect of the orientation due to the flow also come into play^[6].

When the mould is being filled the thermal and mechanical properties of the mould play no part. The mould fills quickly, and the heat from the plastic only penetrates a few millimetres into the wall. The minuscule extra movement of the flowing filling material due to thermal expansion and mechanical forces on the mould is negligible in comparison with the main flow during the fill.

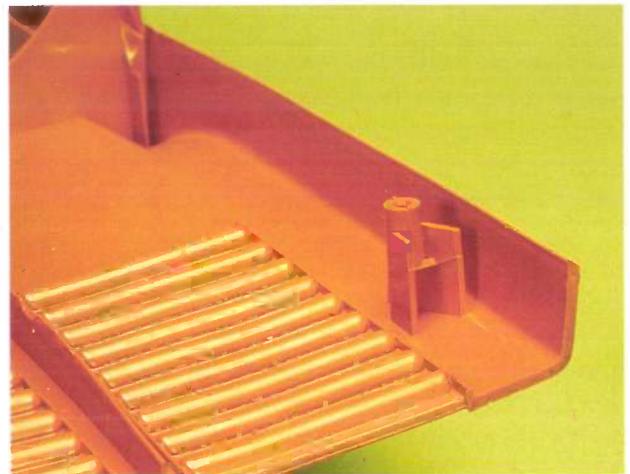


Fig. 7. One exceptionally thin-walled product is the top cover of a vacuum cleaner. (The handle is out of sight, on the left.) The projection on the right of the rim, used for attaching the lower cover, is made hollow to maintain the thin wall. If this was not done there could be undesirable shrinkage on cooling.

The mould designer should however show a keen interest in the thermal and mechanical properties of the mould if he wants to follow the cooling of the product and obtain a better understanding of the final dimensioning of the product. As seen from the solidifying liquid, the pressure dependence of the glass-transition temperature (for amorphous thermoplastics), the viscosity, the viscoelasticity and the coefficient of expansion are important. As seen from the mould, the stiffness, the strength and the thermo-mechanical aspects of the mould and the runner layout are important.

Software relating to the solidifying liquid and developed at Philips Research Laboratories is becoming available (prediction of packing and cooling). For the mould we can use standard finite-element packages such as MARC, ANSYS and ASKA.

When the product is ejected from the mould, it snaps into a shape determined by the equilibrium of the internal forces. This results in shrinkage and possibly warping. Changes in shape due to further cooling can be calculated with new software now being developed at Philips Research Laboratories (prediction of errors in shape and dimensioning^{[3][7]}).

Physical ageing, the gradual recovery of thermodynamic equilibrium in the quenched product, is de-

scribed by a theory that has been developed elsewhere^[8]. For determining the viscoelastic properties of the product use can be made of the correspondence^[9] between linear elasticity and linear viscoelasticity. Standard packages can therefore be used here as well.

[4] C. W. M. Sitters, Numerical simulation of injection moulding, Thesis, Eindhoven 1988.

[5] A. H. M. Boshouwers and J. J. van der Werf, INJECT-3, a simulation code for the filling stage of the injection moulding process of thermoplastics, Thesis, Eindhoven 1988.

[6] G. Eder and H. Janeschitz-Kriegl, Theory of shear-induced crystallization of polymer melts, *Colloid & Polym. Sci.* **266**, 1087-1094, 1988.

[7] A. A. M. Flaman, Het voorspellen van de eigenschappen van gespuitsgiete kunststofprodukten, *Materialen*, No. 2 (February), 40-47, 1988.

[8] L. C. E. Struik, Physical aging in amorphous polymers and other materials, Elsevier, Amsterdam 1978.

[9] F. J. Lockett, Nonlinear viscoelastic solids, Academic Press, London 1972, pp. 32-33.

Summary. This article is based on a talk given at a colloquium held in May 1988 at Philips Research Laboratories. A brief review of a number of plastic products is followed by a short account of the injection-moulding process. This process is analysed as it would be seen by several different specialists: the materials scientist (amorphous plastics, monomers, mixtures, etc.), the product designer (mechanical strength, temperature range, sensitivity to moisture, etc.), the mould designer (dimensions of the cavity in connection with shrinkage, runner systems, separating the mould, etc.) and the rheologist (coefficient of expansion, viscosity, thermal conduction, etc.). The relationship between shaping and the properties of the product is also examined. The article ends with a few words on the help that software can offer in predicting the fill pattern, packing and cooling as well as errors in shape and dimensioning.

A true silicon compiler for the design of complex ICs for digital signal processing

J. L. van Meerbergen and H. De Man

In making software for computers, man has long known that he can save himself much trouble if he uses a relatively easy-to-handle high-level language instead of the very complex internal ones-and-zeros language of the computer. The translation from one language into the other can be done automatically by a dedicated program called a 'compilation program' or a 'compiler'.

In the design of today's integrated circuits the efficient handling of complexity is also a major challenge. Here again a special computer program — or rather a combination of special programs — is found to be the answer to the problem. This approach is often referred to as 'silicon compilation'.

Introduction

The ongoing developments in semiconductor technology, especially in submicron Complementary Metal Oxide Semiconductor (CMOS) processes, will soon permit the integration of millions of transistors on a single integrated circuit (IC or 'chip'). This opens up the possibility of fabricating complete electronic systems on a single chip. There are however a few obstacles to be demolished: despite the immense complexity of the ICs the design should be completed successfully (i.e. without errors) within a reasonably short time (hours or days rather than months or years).

It must also be clear in the earliest phases of the design that some of the specifications for the chip, such as silicon area, power consumption and speed of operation can indeed be met. All this requires massive support for the design process with computers and dedicated software.

In conventional design methods for digital electronic systems most of the time and effort is spent on

translating a given system specification into a 'gate-level' description directly related to the IC implementation, in which the basic descriptive element is the logic gate. The key to a real reduction in design time and design effort is therefore to be found in automation of this first phase in the chip design. Software that can generate the final chip layout starting from a high-level system specification is called a silicon compiler (*fig. 1*). Ideally such a compiler should include tools for architectural synthesis, i.e. it should be able to decide which building blocks (generally called 'modules') are required, how many of each of them and how to interconnect them, to meet a particular system specification. This approach can be called *true* silicon compilation to indicate the difference from certain commercially available silicon compilation software packages which only support the manual composition of 'parametrizable' ^[1] building blocks from a library and their subsequent automatic assembly. In this article we shall describe the silicon compiler PIRAMID (which stands for Precompetitively Inspired Research on Advanced Methods for IC Design). PIRAMID is partly based on the re-

Dr Ir J. L. van Meerbergen is with Philips Research Laboratories, Eindhoven, the Netherlands; Prof. Dr H. De Man is Vice President of IMEC (Interuniversity Micro-Electronics Center) and a professor at the Katholieke Universiteit Leuven, Belgium.

sults of the ESPRIT Project No. 97^[2], more specifically on the software tools for architectural synthesis known as CATHEDRAL II^[3].

In the following sections we shall give a brief account of PIRAMID by describing important subjects such as target architecture, synthesis tools, module generation and 'floorplanning'. Finally, we shall illustrate the operation of our silicon compiler by going step by step through the design of the IC layout of an

ture should not be too restricted, because this would limit the application area of the compiler. The main application of PIRAMID is for ICs for digital signal processing at lower to medium speeds (sampling rates between 1 kHz and 1 MHz) — for speech processing, telecommunications, audio, lower-end image processing etc. This means that ICs must be capable of handling complex matrix-oriented algorithms as well as word-oriented (or even bit-oriented) algorithms. The

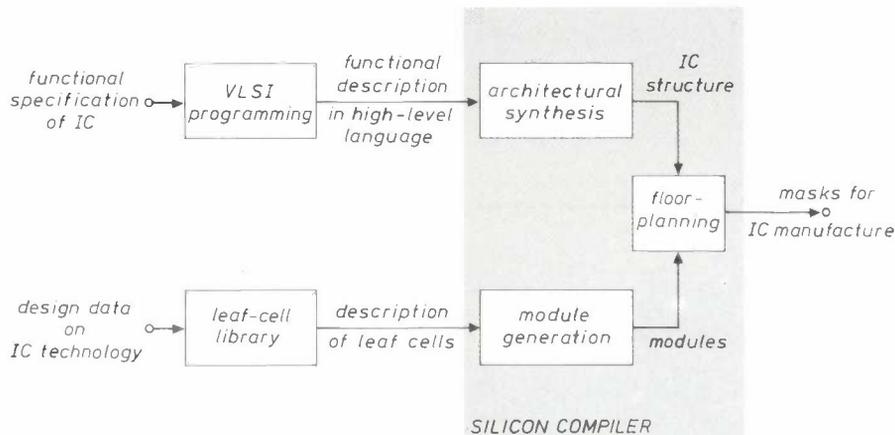


Fig. 1. General diagram for the computer-aided design of complex ICs. The heart of this diagram is the 'silicon compiler', which generates the masks required for the IC manufacture in a highly automated fashion. The starting points for the silicon compiler are a functional description of the IC in a high-level computer language and a description of the basic elements (the 'leaf cells') such as flip-flops from which the IC will be formed. The creation of the high-level functional description of the IC is called *VLSI programming*; this is the responsibility of the system designer. All the necessary information about the IC technology to be used is contained in the descriptions of the leaf cells, which are drawn up by silicon designers. The main components of the silicon compiler are programs for architectural synthesis, module generation and 'floorplanning'. We shall return to these in more detail later.

adaptive digital filter. (As we shall only be considering true silicon compilation we shall drop the adjective 'true' from now on.)

Target architecture of PIRAMID

Silicon compilation requires the definition of the boundaries for the final IC structure and the architectural rules that must always be obeyed. This leads to the notion of a *target architecture* that contains all this information; it gives a fixed framework within which the structure of each particular design can be freely chosen. A target architecture in turn can be defined sensibly only if the intended area of application for the silicon compiler is known. If the target architecture is too general, the result of the silicon compilation can be far from optimum because the specifications are too vague. On the other hand the target architec-

calculations that have to be performed by the ICs range from the vector-product type of calculation (as in convolution, windowing, Fourier transforms and filtering) to nonlinear and data-dependent processing. The target architecture will have to support all of these. The characteristic feature of Digital Signal Processing (DSP) algorithms is that the same processing has to be executed over and over again with new input data. At each execution a fixed and limited period of time — the 'frame period' — is available because the

- [1] 'Parametrizable' means that the value of some of the parameters (such as the wordlength) can be chosen by the designer.
- [2] ESPRIT = European Strategic Programme for Research and development in Information Technology. The name of ESPRIT Project No. 97 is: Advanced algorithms, architecture and layout techniques for VLSI dedicated signal processing chips.
- [3] H. De Man, J. Rabaey, P. Six and L. Claesen, Cathedral II: a silicon compiler for digital signal processing, IEEE Des. & Test Comput. 3, No. 6 (December), 13-25, 1986.

processing must be completed before new input data arrives and the next frame commences. Each frame period consists of a fixed number of execution cycles (or clock cycles). If data-dependent decisions have to be made during the signal processing, we can think of the operations as taking place in a number of parallel operating branches. These branches may have different numbers of execution cycles, and only one branch is selected at each decision. Under all circumstances, however, it is essential that the calculations are completed within the frame period. This means that we are not interested in optimizing the average execution time by bringing in the probability of occurrence of particular branches; we are only interested in the worst-case situation. This is a major difference between real-time DSP and data processing, for example.

In the target architecture of PIRAMID there are three levels of hierarchy^[4]:

- chip level,
- processor level, and
- EXecution Unit (EXU) level.

We shall consider each of these more closely.

Target architecture at chip level

First, at chip level, we can divide up the complete electronic system to be designed into a number of independently operating units ('processors'), dedicated memories for interprocessor communication, input/output units, and a number of buses to connect all these units. Sometimes this breakdown of a system into a number of processors operating in parallel arises quite naturally from the breakdown of the complete block diagram into a number of separate sub-tasks. In other cases it may be less trivial. At present with PIRAMID the breakdown at this level still has to be manual and we shall restrict our attention in this article to single-processor chips, which have target architectures at the two lower levels only.

Target architecture at processor level

At processor level^[5] we can make a distinction between the data path and the controller (or control path); see fig. 2. This is a result of the 'microcoded' approach that we use, as opposed to say a 'hardwired' approach. In our approach the hardware for the actual signal processing (the data path) is separate from the control hardware (the controller). In each clock cycle the controller generates a microcode word that specifies the functions of all parts of the data path in that particular clock cycle. This is a very attractive procedure if the clock rate is ten or more times the frame rate of the signal processing, as in the main application areas envisaged for PIRAMID.

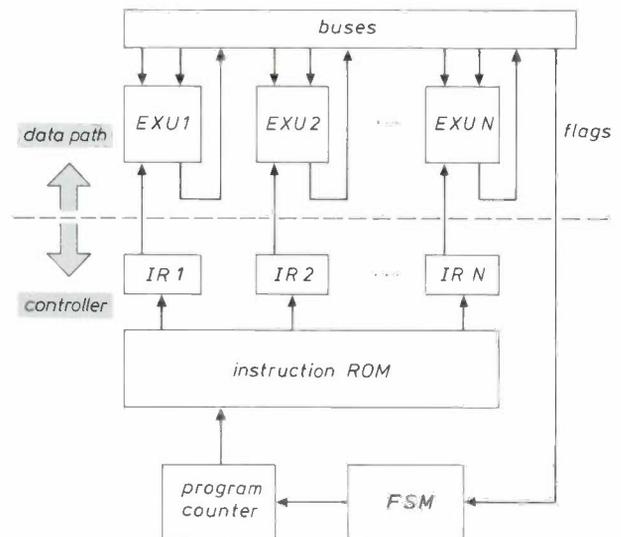


Fig. 2. Target architecture for a single processor in PIRAMID. Each processor designed with this silicon compiler will fit into this general pattern, although there may be considerable differences in factors such as the types and numbers of the EXecution Units (EXUs), the complexity of the buses, the size of the Finite-State Machine (FSM) and the capacity of the instruction ROM. Communications from the controller to the data path go via the Instruction Registers (IRs); there is a separate IR for each EXU. Conversely, the data path can influence the controller by means of 'flags'.

From fig. 2 we see that the data path consists of a number of execution units (EXUs) that cooperate with each other via buses. Typical examples of EXUs are an arithmetic and logic unit (ALU), a random-access memory (RAM), a read-only memory (ROM), an address computation unit (ACU) and a multiplier-accumulator combination (MAC).

The main part of the controller is the instruction ROM that contains all the instructions in the form of microcode words. In each clock cycle the program counter determines the address of the microcode word to be used and the appropriate sections of this word are transmitted to each EXU in the data path via the separate instruction registers IR1, IR2, ... The calculations of the program counter can be influenced by flags indicating the state of the data path (e.g. in the case of data-dependent operations in the program) via the finite-state machine FSM.

Target architecture at EXU level

At the third and lowest level of the architectural hierarchy we have the execution units (EXUs), which are the main products of the module generation (see fig. 1). They are formed from a number of functional building blocks (FBBs) such as adders, decoders, multiplexers, register files and output bus drivers (see fig. 3). The heart of each EXU is an Operation Unit (OPU), like a multiplier, a RAM, a ROM and so on,

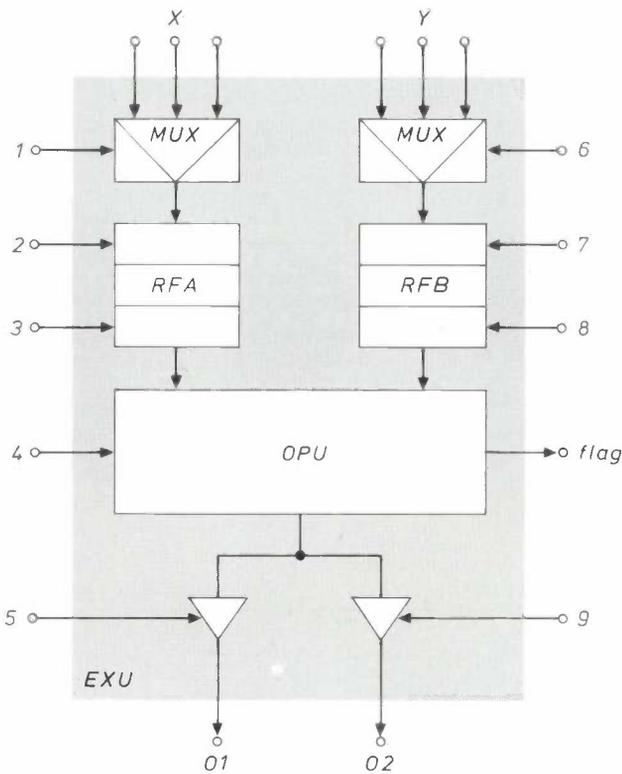


Fig. 3. General target architecture of an EXecution Unit (EXU) in PIRAMID. At the centre is the OPeration Unit (OPU) that determines the type of EXU we have. The OPU has two inputs and one output; at each input there is a register file RF. The data words for the register files are selected by the multiplexers MUX from a number of input buses X,Y. The output of the OPU is fed to the bus terminals O1 and O2 via two bus drivers. The terminals 1, 2, . . . , 9 are connected to the instruction registers and the terminal 'flag' is connected to the FSM of the controller (see fig. 2).

which actually determines the type of the EXU. An OPU has two inputs and one output. At each input a register file is available. This model applies not only for arithmetic units, but also for memories which have two inputs (one for the data and one for the address) and one output (for data). In PIRAMID the FBBs have parameters such as the wordlengths, the number of registers in the register files and the type of adder. For any particular use of the silicon compiler these parameters are chosen optimally and the FBBs are accordingly made 'to measure' from the more elementary leaf cells stored in the leaf-cell library.

The PIRAMID silicon compiler

A schematic representation of the PIRAMID silicon compiler is shown in fig. 4. The three main components (often called 'environments') are the same as in the general silicon compiler of fig. 1:

- the *architectural synthesis environment* (based on CATHEDRAL II)^[3], in which a high-level functional

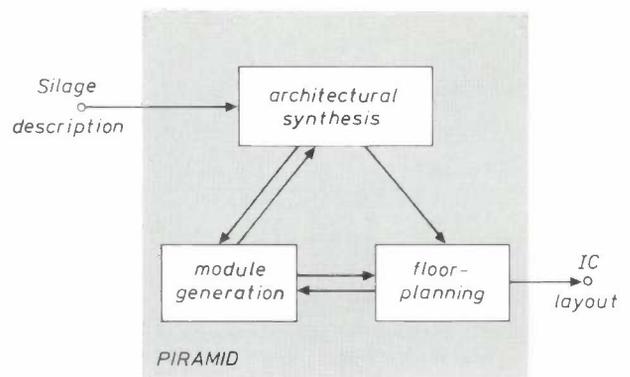


Fig. 4. Schematic representation of the PIRAMID silicon compiler. The complete software package, which translates a high-level chip specification (in the computer language 'Silage') into a detailed chip layout, has three components: architectural synthesis, module generation and floorplanning. The most important data flows during the design of the layout are indicated by arrows.

description of the IC is translated into a specific structure that fits within the target architecture at all levels;

- the *floorplanning environment*, which translates this specific structure into a fully detailed layout of the final chip;

- the *module-generation environment*, which supplies the other two environments with specific information ('views') of any module for specified parameter values. There are many different views, such as a black-box view, a detailed layout view, an area view, a 'timing' view, a dissipation view, a test view, a functional view or a structural view.

A characteristic feature of PIRAMID is the close interaction between architectural synthesis and module generation and between module generation and floorplanning. As the user of PIRAMID is not usually concerned with the description of leaf cells, these are taken for granted and do not occur in the scheme of fig. 4. In the following sections we shall consider the three environments of PIRAMID in more detail and we shall try to illustrate some of their key features by following the design of a particular digital-signal-processing IC all the way through the PIRAMID silicon-compilation process.

Example: the design of an adaptive digital filter

Fig. 5 shows the block diagram of an adaptive digital filter for which we want to use PIRAMID to design an IC layout. In this filter with time-dependent

[4] F. Catthoor *et al.*, Architectural strategies for an application-specific synchronous multiprocessor environment, IEEE Trans. ASSP-36, 265-284, 1988.
 [5] See also: J. L. van Meerbergen, Developments in integrated digital signal processors, and the PCB 5010, Philips Tech. Rev. 44, 1-14, 1988. (Also reprinted as chapter 15 in [7].)

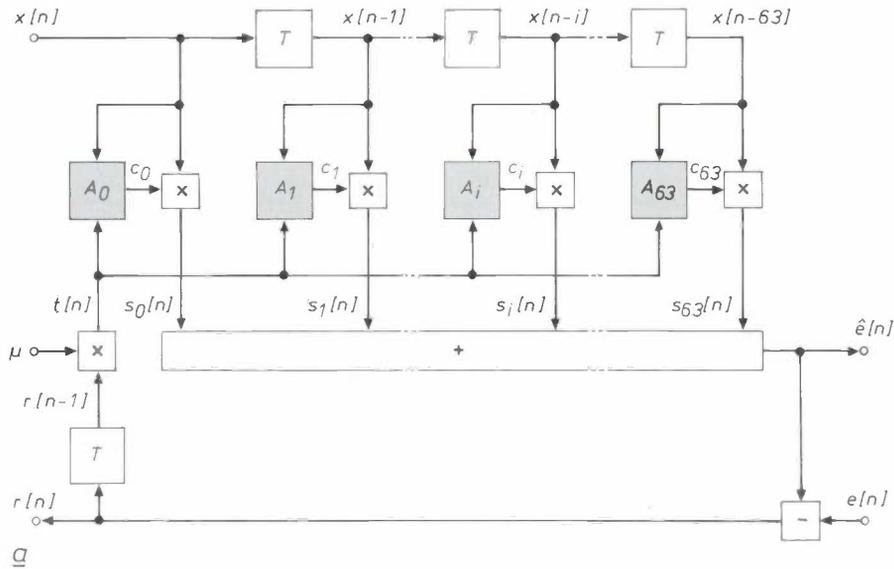
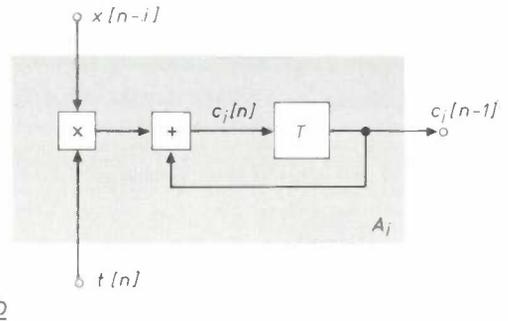


Fig. 5. a) Block diagram of a transversal adaptive digital filter. The input signal $x[n]$ and delayed versions $x[n-1], \dots, x[n-63]$ are multiplied by the filter coefficients c_0, c_1, \dots, c_{63} and summed to produce the output signal $\hat{e}[n]$. This is compared with a reference signal $e[n]$. The filter coefficients are automatically adapted so as to minimize the difference signal $r[n]$. The calculations required for the adaptation are made in the adaptation units A_0, A_1, \dots, A_{63} . b) Block diagram of the adaptation unit A_i . All the operations indicated in this figure have to be executed once during each sampling interval T of the input signal $x[n]$. (μ = adaptation constant.)



filter coefficients c_0, c_1, \dots, c_{63} the input signal $x[n]$ is filtered. The output signal is $\hat{e}[n]$. The filter coefficients are adapted automatically so that $\hat{e}[n]$ resembles a given reference signal $e[n]$ as closely as possible. Depending on the particular application the actual output of interest from the adaptive filter can be either the reconstructed signal $\hat{e}[n]$, the difference signal $r[n] = e[n] - \hat{e}[n]$ or the set of filter coefficients c_0, c_1, \dots, c_{63} [6]. For present purposes all we need to say is that adaptive filters automatically modify their frequency characteristics in accordance with some predefined criterion and that they can be used in many applications: automatic echo cancellation, automatic cable equalization and so on [7].

Before we can start the silicon compilation, the block diagram must be described in a high-level computer language. We have chosen the language Silage [8]. Silage is particularly useful for describing signal-processing algorithms. A Silage program can be looked upon as a textual representation of a signal-flow diagram. It consists of 'definitions' or 'equations' rather than 'assignments'. The basic objects in Silage represent signals, and each signal is only defined once in the program. This means that the

order of the definitions is of no significance. The complete Silage description of the filter of fig. 5 is given in fig. 6.

An essential feature of Silage is the use of 'types'. The type WORD defined by 'num < 32, 12 >' means that we are working with a wordlength of 32 bits, with

```

#define mu 0.1
#define WORD num<32,12>

func main (input, e : WORD) r % WORD =
begin
  sum[0]=WORD(0);
  x = input;
  t = WORD(r@1*WORD(mu))
  ( i : 0..63)::
    begin
      c[i] = c[i]@1 + WORD(t*x@i);
      s[i] = WORD(x@i*c[i]@1);
      sum[i+1]=sum[i]+s[i];
    end;
  ehat = sum[64];
  r = e - ehat;
end;
    
```

Fig. 6. Complete Silage description of the adaptive digital filter of fig. 5 (without 'pragmas').

12 bits after the decimal point (it is really a binary point, of course). It is a very convenient way of implementing the wordlength specifications that occur frequently in digital signal processing. Another feature of Silage is the use of '@i' to represent a delay iT corresponding to i sampling intervals (see fig. 5). Square brackets are used in Silage as a short-hand notation for variables used in an iteration.

Now that we have obtained a Silage description, we can commence the actual silicon compilation, starting with the architectural synthesis.

Architectural synthesis

In the first phase of the silicon-compilation process the structure of the data path and the structure of the controller are determined. Parameter values (such as wordlengths, memory capacities and numbers of registers in a register file) are chosen. The microcode for the controller is also determined, so that the contents of the instruction ROM are defined. All this is done interactively, which means that the designer can easily iterate the synthesis process from his workstation until he is completely satisfied with the results. During the interaction the module-generation environment is consulted in depth to give early information about the expected chip area and timing performance. Finally, the result of the synthesis is sent to the floorplanning environment.

A more detailed picture of the synthesis process, which was based on CATHEDRAL II [3], is given in fig. 7. First the Silage description is processed by the program JACK [9]. This generates a structure for the data path by selecting appropriate EXUs with the help of a 'rule base'. This is a knowledge base stored in the computer and put together by experienced designers, who have identified a set of priority rules for mapping Silage operations on to the various available EXUs.

The structural information of the data path produced by JACK can also be represented graphically. For the adaptive digital filter the result is the block diagram shown in fig. 8.

JACK also translates the Silage description of a system into a set of register-transfer equations of the form:

$$\text{reg_3: var_x} \leftarrow \text{reg_2: var_y, reg_3: var_z} \mid \text{alu} = \text{SUB.}$$

This *register-transfer language* describes the relationship between pairs of variables and the register files in which they are stored. So the equation above means that variable z, stored in register file 3, must be subtracted from variable y in register file 2. The result is called variable x and is stored in register file 3. (The real names of the register files are more complicated

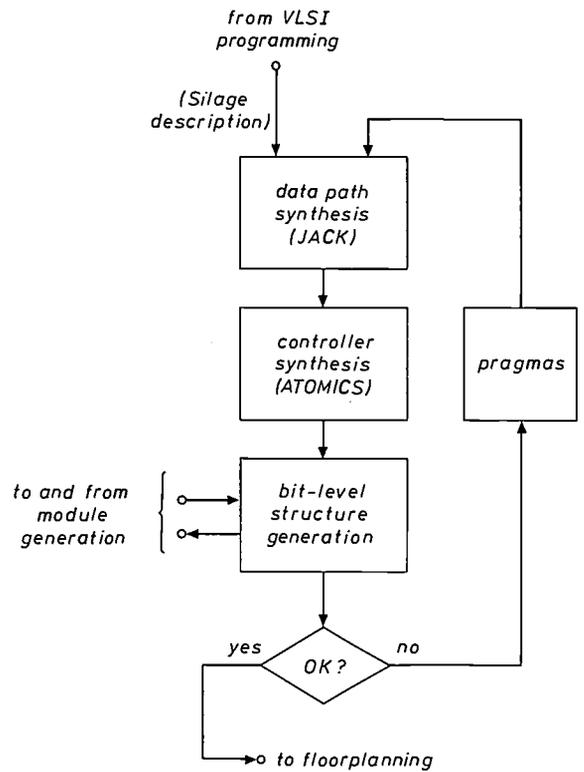


Fig. 7. Details of the architectural synthesis environment of PIRAMID. Starting from the Silage description of a chip, a data path is first synthesized by the computer program JACK. Then the controller is composed by the program ATOMICS, which is called a 'microcode scheduler'. Finally, the structure of the chip is generated at bit level and information is called from the module-generation environment. If the designer is satisfied with the results, floorplanning of the chip can start; if not, the synthesis phase is reiterated with the inclusion of 'pragmas', which 'steer' the synthesis process in some preferred direction.

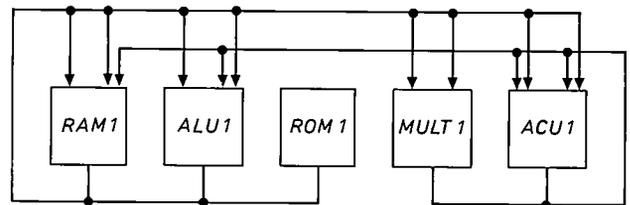


Fig. 8. Block diagram of the data path of the adaptive digital filter as synthesized by JACK. It consists of five execution units: a random-access memory RAM1, an arithmetic and logic unit ALU1, a read-only memory ROM1, a multiplier MULT1 and an address-computation unit ACU1 for RAM1.

[6] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, Adaptive techniques for signal processing in communications, IEEE Commun. Mag. 23, No. 11 (November), 8-19, 1985.
 [7] A. W. M. van den Eenden and N. A. M. Verhoeckx, Discrete-time signal processing — an introduction, Prentice Hall, New York 1989.
 [8] P. N. Hillfinger, A high-level language and silicon compiler for digital signal processing, Proc. IEEE Custom Integrated Circuits Conf., Portland, Oreg., 1985, pp. 213-216.
 [9] J. Vanhoof, J. Rabaey and H. De Man, A knowledge-based CAD system for synthesis of multi-processor digital signal processing chips, Proc. IFIP-TC10/WG10.5 Int. Conf. on Very Large Scale Integration, Vancouver 1987, pp. 73-88.

because the name of the particular EXU it belongs to is included in the name, e.g. *reg_1_alu_4*.)

The output of JACK forms the input to the next computer program, the *microcode scheduler* ATOMICS^[10], whose main task is the synthesis of the controller. The first task for ATOMICS is to put the set of register-transfer equations *in the right sequence* in terms of clock cycles and to determine which transfers can be made simultaneously (i.e. in the same clock cycle). The total number of clock cycles in one frame period is also calculated. For our present design this total number is 4 times the number of taps of the adaptive filter, i.e. $4 \times 64 = 256$ clock cycles. However, the minimum duration of the frame period (and its reciprocal, the maximum operating rate of the chip) will be known only after we have calculated the minimum duration of a clock cycle for the 'slowest' of all the EXUs on the chip. This will be done later on.

The next task for ATOMICS is to associate each register transfer with particular registers in the register files of the EXUs. ATOMICS tries to minimize the number of registers by performing a lifetime analysis for each variable. Variables whose lifetimes do not overlap can go into the same register (*fig. 9*).

Register file	Register No.	Variable	Lifetime intervals
reg_1_alu_1	0	e	[17, 19]
	1	sum	[9, 16]
		c	[12, 13]

Fig. 9. Example of the lifetime analysis by ATOMICS for the dimensioning of register files. In register file 1 of ALU1 of the adaptive digital filter three variables *e*, *sum* and *c* have to be stored during certain time intervals in the signal processing. Since ATOMICS finds that during each frame period *e* occurs from clock cycle 17 until clock cycle 19 and *sum* occurs from clock cycle 9 until clock cycle 16, *e* and *sum* can be stored in the same register. Therefore the register file has to contain only two registers.

The third task for ATOMICS is to generate a symbolic microcode. It is called 'symbolic' because it consists of the kinds of symbols used in assembler languages (see *fig. 10*), instead of the almost unreadable strings of ones and zeros used in machine code. Afterwards the symbolic microcode can be converted automatically into 'true' microcode, i.e. into machine code.

A detailed structure of the chip at bit level can now be generated, i.e. for each EXU, FBB, OPU and bus the dimensions can be determined in numbers of bits. This is done with the set of programs labelled *bit-level structure generation* in *fig. 7*. These programs combine the outputs from JACK and ATOMICS with information called from the module-generation environ-

ment. As a first action the set of parameter values of each individual EXU is defined; these parameter values are used to call the corresponding area and timing views. In this way the designer can obtain detailed information about each EXU in the block diagram

```
reg_1_alu_1:W[0],
buf_1_alu_1:write,
reg_2_alu_1:R[0],
alu_1:add
reg_1_alu_1:R[0],
```

Fig. 10. Example of one register transfer in symbolic microcode, as supplied by ATOMICS. The complete microcode for the adaptive digital filter consists of 23 such transfers. After conversion into 'true' microcode we obtain 23 microcode words, each of 70 bits, which will be stored in an instruction ROM with a capacity of 23×70 bits.

```
DESIGN adapt
MODULE mult_1 mult
  PARM
    NX = 1: MINIMUM = 0, MAXIMUM = 3;
    NY = 1: MINIMUM = 0, MAXIMUM = 3;
    X  = 32: MINIMUM = 0;
    Y  = 32: MINIMUM = 0;
    O1 = 64: MINIMUM = 0;
    O2 = 0: MINIMUM = 0;
    RX = 2: MINIMUM = 0, MAXIMUM = 16;
    RY = 1: MINIMUM = 0, MAXIMUM = 16;
    ACC = 0: MINIMUM = 0, MAXIMUM = 1;
  ENDPARM
ENDMODULE
ENDDESIGN
```

a

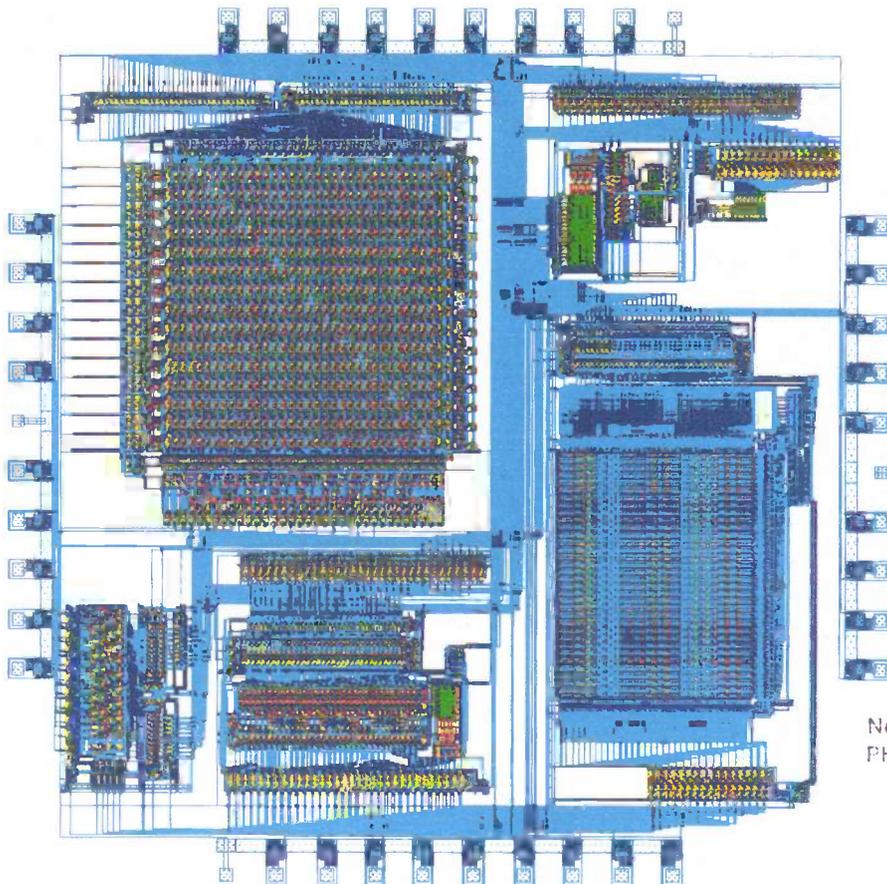
EXU	SIZE (bits)	R1	R2	TIMING (ns)	AREA (mm*mm)
ram_1	132*32	1	2	52	2.3*3.7= 8.5
mult_1	32*32->64	2	1	214	2.9*3.8=11.2
alu_1	32	2	1	79	1.8*1.7= 3.2
rom_1	8*32	-	-	20	0.2*0.6= 0.1
acu_1	8	3	2	74	1.6*1.1= 1.9
ctrl rom	23*70	-	-	26	1.0*0.4= 0.4
					----- +
					25.3

b

Fig. 11. *a*) Parameter file describing the 32-bit \times 32-bit multiplier of the adaptive digital filter. (See also the general target architecture for EXUs, *fig. 3*). *NX*, *NY* number of inputs to the input multiplexers; *X*, *Y* wordlength of the inputs to OPU; *O1*, *O2* wordlength of the outputs of OPU; *O2* = 0 indicates that only one of the outputs of OPU is used; *RX*, *RY* number of registers in the register files; *ACC* = 0 indicates that no accumulator is included in this EXU. *b*) Architectural summary of the adaptive digital filter. For each EXU of *fig. 8* the most important characteristics are listed here. The first column specifies the type of EXU; the second column lists the size in bits. *R1* and *R2* give the number of registers in each register file. The column *Timing* is of considerable significance: it shows the time (in ns) required for one complete operation by each of the EXUs; the multiplier requires the longest time (214 ns). This will be the limiting factor for the operating rate of the complete chip. The column *Area* shows the net area of each EXU in mm^2 . Together with the instruction ROM they add up to 25.3 mm^2 .

and the main characteristics of the complete IC. By way of illustration *fig. 11a* shows the parameter file for the 32-bit \times 32-bit multiplier and *fig. 11b* summarizes the architecture of our adaptive digital filter. From the data of *fig. 11b* the designer can calculate the maximum operating rate f_{op} of the chip design. The multiplier limits the minimum clock cycle time to 214 ns. As we need 256 clock cycles in one frame period

processing. Or he might want to reduce the number of EXUs if there is speed to spare. It is also possible to steer the synthesis process in some preferred direction by assigning certain operations to a particular EXU, for example. Here the designer can use *pragmas* — our name for ‘hints’ for the synthesis programs. Later on we shall give examples of results obtained by using *pragmas*.



Ned. Philips Bedrijven B.V.
PHILIPS RESEARCH LABS.
LIBRARY WY - 1
P.O. Box 80 000
5600 JA EINDHOVEN
THE NETHERLANDS

Fig. 12. Layout of the adaptive digital filter of *fig. 5*, obtained by highly automated design with the PIRAMID silicon compiler. The large block at the top left is the multiplier; the block at the lower right is the RAM. The rest of the data path is at the lower left; the controller is at the top right. The total area amounts to 59 mm² if the IC is fabricated in 1.6- μ m CMOS technology.

for this design we get $f_{op} = 1/(256 \times 214 \times 10^{-9}) \approx 18$ kHz.

As a second action in this phase of the silicon compilation buses are merged wherever possible to minimize the chip area required for the interconnections.

Armed with the information given in *fig. 11* the designer can decide to start the floorplanning, or reiterate the synthesis procedure. If the result of the synthesis does not meet the design specifications for speed, he can try to add more EXUs to speed up the

Floorplanning

When the designer is satisfied with the results of the architectural synthesis, the floorplanning — i.e. the definitive shaping of the IC layout — can start. The floorplanning is done in several steps. First, black-box views of the modules selected during the synthesis

^[10] G. Goossens, J. Rabaey, J. Vandewalle and H. De Man, An efficient microcode-compiler for custom DSP-processors, Proc. IEEE Int. Conf. on Computer-Aided Design, Santa Clara, Cal., 1987, pp. 24-27.

are called from the module-generation environment. These are automatically located to give a compact plan with an approximately rectangular outline. In this process modules can be reflected, rotated through 90°, 180° or 270° and the aspect ratio of some modules can be changed.

A Bus Control Block (BCB) is added to every bus, and all the BCBs are connected in series by scan lines, which are used for testing the ICs after manufacture.

Next, the optimum positions for the bonding pads are determined; these form the connections of the IC to the outside world. Different chip boundary cells, each comprising one bonding pad, are available and can be called from the module-generation environment.

The most important task remaining is the exact determination of the shape and size of the multilayer interconnections (the 'internal wiring' of the chip) to be made in the spaces ('channels') between the various modules. This is called *routing* and consists of two phases. In the first phase we have *global routing* — determining which channels are used to make each interconnection. In the second phase the *channel routing* program completes this part of the design by supplying the details for each channel: all the interconnections are now accurately positioned with the required widths, distances, etc.

Once the routing has been completed, the module-generation environment is called again for detailed layout views of all the modules. A fully detailed layout database is then available and the masks for the IC fabrication can be drawn.

Fig. 12 shows the layout of the adaptive digital filter of fig. 5 as derived by our highly automated computer method. It has a total area of 59 mm². The main modules of the system can easily be recognized. The controller situated at the top right has also been produced by automated design in accordance with the target architecture of fig. 2.

Module generation

As the preceding sections have shown, the module-generation environment provides a most important support for the other two basic components of PIRAMID (synthesis and floorplanning). It is very largely this environment that determines the versatility and efficiency (e.g. the speed) of the overall silicon compilation. This environment should 'instantly' make the right view available for every possible module at the parameter values specified by the other two environments.

The module-generation environment has its own tools for placement, routing and abutment of leaf cells or functional building blocks within the modules.

(Abutment means that interconnections between cells arise 'naturally' when they are neighbours.)

In the module generation a special computer language, Modgen, is used. An example is given in fig. 13, where an imaginary n -bit execution unit AMC (Adder/Multiplexer Combination) is described in Modgen with n as a parameter.

The creation of the module-generation environment has demanded considerable design effort, as can be

```

cell admux(n)
int n
{
  place fa[0] = fulladder;
  place mux[0].ll = multiplexer at fa[0].ul;
  for ( i=1; i<n; i++) {
    place fa[i].ll = fulladder at fa[i-1].lr;
    place mux[i].ll = multiplexer at mux[i-1].lr;
  }
  routerbox fa_to_mux {
    for ( i = 0; i<n-1; i++)
      net fa_out[i] = mux[i].in, fa[i].out
  }
  route fa_to_mux gcr;
}

```

Fig. 13. Example of the computer language Modgen used in module generation. An imaginary n -bit execution unit AMC (Adder/Multiplexer Combination) is described here in Modgen. The parameter n can be specified by the user. Two types of leaf cells are used: a 1-bit full-adder cell and a 1-bit multiplexer cell. The constructs that add topological information such as the placement of cells next to each other (see lines starting with 'place') are characteristic of Modgen. Connections can be made automatically by abutment if the terminal positions of two adjacent cells match. Another way of making interconnections between cells is to use router software, as shown in the last few lines of the example. Use is made here of GCR, a routing method originating from the University of Delft.

seen from the summary of the current module library (fig. 14). It is therefore very important that it should not become obsolete as soon as a new IC technology is introduced. A new technology will usually require updating of the leaf-cell library and the module library. In our experience the module-generation environment of PIRAMID can adequately be modified to be compatible with new design rules resulting from changes in IC technology^[11].

Alternative designs

The complete silicon-compilation process for the IC design of our adaptive digital filter, from the Silage description of fig. 6 to the layout drawing of the layout of fig. 12, took only a few hours on an Apollo DN3000 workstation. If the designer had not been happy with the values shown in fig. 11b after the first

MODULE	DESIGN TIME (manmonths)	NUMBER OF LEAF CELLS	SOURCE-CODE (No. of lines)
Address Calculation Unit (ACU)	12	19	375
Arithmetic and Logic Unit (ALU)	16	68	1400
Bus Control Block (BCB)	2	4	52
Multiplier ACcumulator (MAC)	18	87	2800
Programmable Logic Array (PLA)	3	17	163
Programmable Logic Unit (PLU)	6	7	370
Random-Access Memory (RAM)	24	60	5170
REGister file (REG)	12	20	1070
Read-Only Memory (ROM)	12	29	420
TRIstate buffer (TRI)	1	8	630
	— +		— +
	106		12450

Fig. 14. At present the module library of the module-generation environment in PIRAMID contains the software descriptions of 10 different parametrizable modules. To give an impression of the complexity of this software, three characteristics of the modules are listed here: the design time in man-months, the number of different leaf cells used and the number of program lines (in the computer language Modgen).

architectural-synthesis phase, he could easily have made a completely different design with the aid of pragmas. We should now like to illustrate this for two alternative designs.

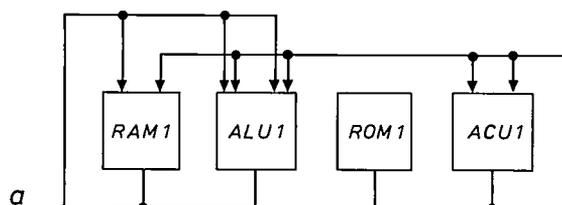
Alternative design I

This design starts by including a pragma that states that no EXU of the multiplier type should occur. This implies that all multiplications in the signal processing on the chip must be made in an EXU of the arithmetic and logic unit type, by multiple shifting and adding. We can expect that in this way the chip area and also the maximum operating rate will decrease. The block diagram resulting from this design is shown in fig. 15, with a table that summarizes the architecture. The combined net area of all EXUs plus instruction ROM amounts to 15.9 mm² and the minimum clock-cycle time is now 120 ns (determined by the ALU). Because we also know from the output of ATOMICS that each frame period must contain 64 × 35 cycles, the minimum frame period for this adaptive digital filter is 120 × 64 × 35 × 10⁻⁹ = 268 μs. This corresponds to a maximum operating rate of 3.7 kHz.

A detailed IC layout resulting from this design is shown in fig. 16b. The total area of this chip is 44 mm².

Alternative design II

If the design of the adaptive digital filter is repeated once more, but now with a pragma that states that one multiplier EXU should be used and two RAMs instead of one (as in the original design), the block dia-



EXU	SIZE (bits)	R1	R2	TIMING (ns)	AREA (mm*mm)
ram_1	132*32	1	4	52	2.3*3.7= 8.5
alu_1	32	5	5	120	1.8*2.6= 4.7
rom_1	7*32	-	-	20	0.6*0.2= 0.1
acu_1	8	5	2	74	1.6*1.1= 1.8
ctrl rom	67*73	-	-	30	1.1*0.7= 0.8
					— +
					15.9

Fig. 15. Alternative design I. a) Block diagram. The design was started with the pragma that no EXU of the multiplier type should be used. This reduces the chip area, but also the maximum operating rate of the chip. b) Architectural summary.

[11] The CD90 (CAD and Design for the Nineties) project, based on the principles discussed in this article, has been started in the IC Design Centre at Philips Research Laboratories in Eindhoven, the Netherlands. The aim of the project is to bring the basic ideas and prototype tools for silicon compilation to industrial maturity. The project includes 'full testability' features and the fabrication of a prototype IC containing all the electronic functions for Compact Disc Digital Audio on a single chip. The CD90 project will be described in a forthcoming issue of Philips Technical Review.

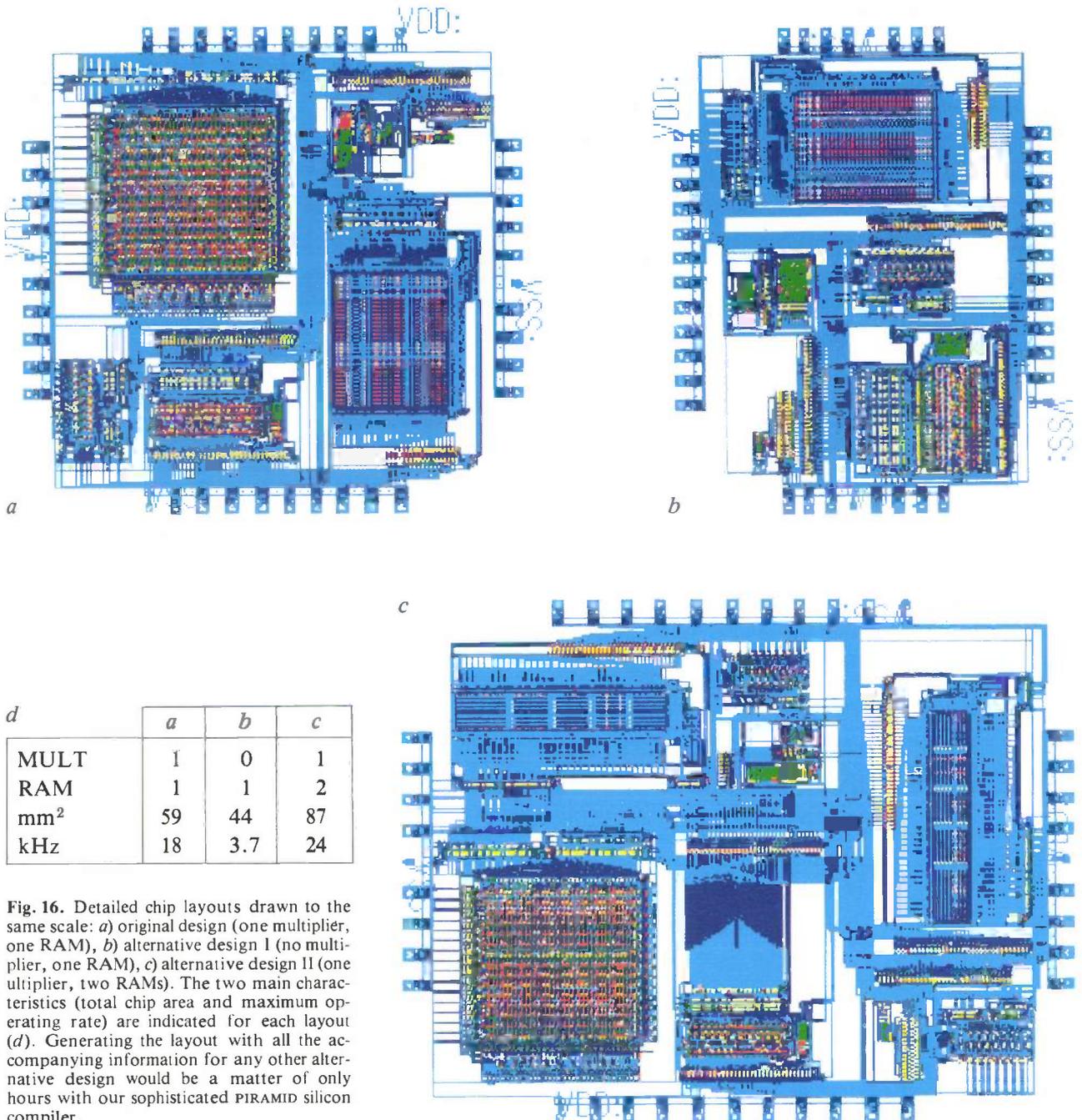


Fig. 16. Detailed chip layouts drawn to the same scale: *a*) original design (one multiplier, one RAM), *b*) alternative design I (no multiplier, one RAM), *c*) alternative design II (one multiplier, two RAMs). The two main characteristics (total chip area and maximum operating rate) are indicated for each layout (*d*). Generating the layout with all the accompanying information for any other alternative design would be a matter of only hours with our sophisticated PIRAMID silicon compiler.

gram and the architectural summary of *fig. 17* result. The corresponding IC layout is shown in *fig. 16c*. From the output of ATOMICS we can learn that for this design only $64 \times 3 = 192$ clock cycles are needed in each frame period. As the minimum clock cycle time is again 214 ns (as in the original design) the maximum operating rate of this chip is: $f_{op} = 1/(192 \times 214 \times 10^{-9}) \approx 24$ kHz. From *fig. 17b* we see that the combined net area of all EXU's plus instruction ROM is now 30.5 mm². After completion of the floorplanning we find a total area of 87 mm².

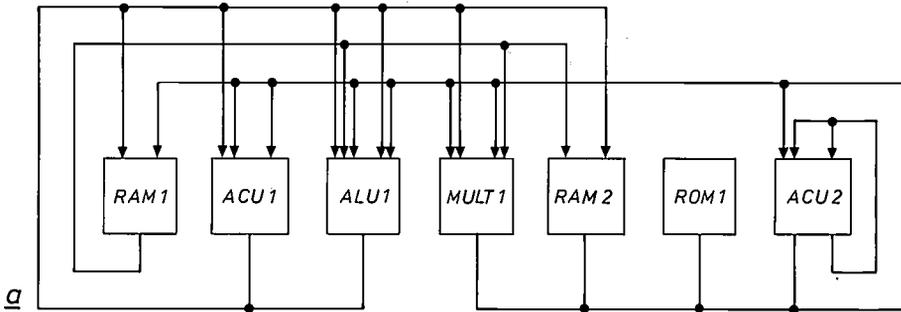
Simulation and testing

So far we have tacitly assumed that we are in an ideal world: nothing goes wrong in the design process and no errors at all are made. But how can we be sure that the IC layout really represents the system we want to design? In fact we do *not* know and we have to carefully verify the intermediate results obtained at various stages in the design process. We use *computer simulations* to check whether the system description we use at a particular stage still corresponds to our original system definitions.

In PIRAMID there are two points that are particularly suitable for computer simulations. The first is just before the start of the architectural synthesis. As the Silage description is an *executable specification* it can be used directly for simulation. We can do this with the computer-simulation program YASS [12], which checks the system description before we start any compilation. The second point is just after the synthesis. Here the functional descriptions of the controller

rect operation of the individual chips will always require some checking.

With complex chips efficient testing is only feasible if test options have been built in during the early stages of the design. In the section on floorplanning we have already mentioned in passing the existence of Bus Control Blocks (BCBs) for test purposes. We shall not pursue the matter here, but it is clear that 'design-for-testability' deserves serious attention [14].



EXU	SIZE (bits)	R1	R2	TIMING (ns)	AREA (mm*mm)
ram_1	67*32	1	1	52	2.3*2.7= 6.1
ram_2	65*32	1	1	52	2.3*2.7= 6.1
mult_1	32*32->64	2	1	214	2.9*4.0=11.6
alu_1	32	2	1	79	1.8*1.7= 3.2
rom_1	7*32	-	-	20	0.6*0.2= 0.1
acu_2	7	1	1	68	1.4*1.0= 1.4
acu_1	7	2	2	68	1.5*1.1= 1.6
ctrl rom	18*90	-	-	28	1.2*0.3= 0.4
					----- +
					30.5

Concluding remarks

The approach used in the design of the PIRAMID silicon compiler has been found very effective: the integration of architectural synthesis (based on CATHEDRAL II), module generation and floorplanning, in parallel with the development of a software leaf-cell library, yields a very powerful system for the design of complex digital ICs. It has been found that design with PIRAMID is much faster than with conventional methods (hours or days rather than months or years).

The advantages of PIRAMID are the result of our 'architecture-driven' approach to the problem of silicon compilation. To elaborate, the most essential features of PIRAMID are:

- Accurate data ('views') for timing and area at the EXU level are available early on in the compilation process.
- A large 'design space' can be explored since it is easy to evaluate different 'solutions' for the architectural synthesis.
- Our 'pragmas' give flexible and effective interaction between designer and silicon compiler.

Fig. 17. Alternative design II. *a)* Block diagram. This design was started with the pragmas that one multiplier EXU, two RAM EXUs and two ACU EXUs should be used. This increases the maximum operating rate at the price of a larger chip area. *b)* Architectural summary.

and of all the EXUs in the data path are available in a 'high-level hardware-description language'. These descriptions are used in the simulation program HELIX [13]. The designer can compare the results of this simulation with those of the earlier YASS simulation and with the original system definition. If no discrepancies are found, he can be confident that the layout produced by the floorplanning will indeed represent the system he wished to design.

Between the IC-mask data base generated by PIRAMID and the actual hardware of the chip there is the long and complicated IC fabrication process. Even if the masks are perfect, there will never be a 100% yield from fabrication. This means that the cor-

[12] F. D. Schlij, private communication; see also the article referred to in [11].

[13] HELIX is a trademark of Silvar Lisco.

[14] F. P. M. Beenker, K. J. E. van Eerdewijk, R. B. W. Gerritsen, F. N. Peacock and M. van der Star, Macro-testing: unifying IC and board test, IEEE Des. & Test Comput. 3, No. 6 (December), 26-32, 1986.

• New EXUs can easily be incorporated in the existing module generation and rule-based synthesis of PIRAMID. The adaptive digital filter discussed in this article is only one of many successful applications of PIRAMID. It has also been tried out for a wide range of other chip designs in 1.6- μm IC technology, such as a vocoder, a quadrature mirror filter, a wideband speech codec, wave digital filters and — to quote a quite different example — in a syndrome-generation circuit for error correction^[15].

Acknowledgments

The silicon compiler described here is the result of fruitful cooperation between scientists from IMEC (Interuniversity Micro-Electronics Center, Leuven, Belgium) and scientists from Philips Research Laboratories (Eindhoven, the Netherlands). The original ideas for the target architecture came about through cooperation within ESPRIT (the European Strategic Programme for Research and development in Information Technology). Prototype software for the CATHEDRAL II synthesis tools has been written by the VSLI Systems Design Group at IMEC.

Philips Research designed the actual silicon compiler PIRAMID, including the rule base and the module library with its development tools. Contributions from the University of Delft and from Philips Telecommunication and Data Systems on placement and routing are also gratefully acknowledged.

Summary. PIRAMID is a sophisticated software package ('silicon compiler') for the highly automated design of complex integrated circuits (ICs) for digital signal processing in areas such as speech processing, telecommunications, audio and lower-end image processing. A general target architecture for IC design work has been defined, with a data path consisting of a number of execution units and a controller which controls the data path via a microcode. PIRAMID consists of three environments: architectural synthesis (based on CATHEDRAL II computer programs), floorplanning and module generation. The silicon compilation starts from a system description in the high-level computer language Silage and a library of elementary units ('leaf cells') that contain all the relevant design information about the IC technology to be used; it finishes with a fully detailed IC-layout database. The module-generation environment makes use of the leaf cells to design the execution units and the controller, with the parameter values (such as wordlengths, memory capacities, number of registers in a register file) specified by the other two environments. The designer can 'steer' the compilation process by giving specific hints ('pragmas'). The main features of PIRAMID are illustrated in this article by tracing through the different stages in the IC design for a 64-tap adaptive digital filter.

^[15] A. Delaruelle, Design of a syndrome generator chip using the PIRAMID design system, Proc. Eur. Solid-State Circuits Conf., Manchester 1988, pp. 256-259.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale, P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

R. A. Stolzenberger	N	Nonlinear optical properties of flux growth KTiOPO ₄	Appl. Opt. 27	3883-3886	1988
G. M. Loiacono & R. A. Stolzenberger	N	Observation of complex domain walls in KTiOPO ₄	Appl. Phys. Lett. 53	1498-1499	1988
J. B. Oude Elferink*, F. H. P. M. Habraken*, W. F. van der Weg* (* Univ. Utrecht) & A. E. T. Kuiper	E	Hydrogen in oxidized silicon oxynitride thin films	Appl. Surf. Sci. 33-34	765-772	1988
J. Ühlken*, R. Waser & H. Wiese* (* Tech. Hochschule, Darmstadt)	A	A Fourier transform impedance spectrometer using logarithmically spaced time samples	Ber. Bunsenges. Phys. Chem. 92	730-735	1988
H. Boulard & Y. Kamp	B	Auto-association by multilayer perceptrons and singular value decomposition	Biol. Cybern. 59	291-294	1988
H. Greiner	A	The Gibbs energy of a chemical reaction system considered as a function of its elemental abundancies	CALPHAD 12	143-154	1988
H. Greiner	A	The chemical equilibrium problem for a multiphase system formulated as a convex program	<i>ibid.</i>	155-170	1988
F. P. Biemans & C. A. Vissers (Univ. of Technol., Enschede)	N	Computational tasks in robotics and factory automation	Comput. Ind. 10	95-112	1988
C. Ronse	B	Minimal test patterns for connectivity preservation in parallel thinning algorithms for binary digital images	Discrete Appl. Math. 21	67-79	1988
P. J. A. Thijs, J. M. Lagemaat & R. Woltjer	E	High mobility two-dimensional electron gas in InP/Ga _{0.47} In _{0.53} As heterojunctions grown by low-pressure organometallic vapour phase epitaxy	Electron. Lett. 24	226-227	1988
A. Valster, J. P. André, E. Dupont-Nivet & G. M. Martin	E,L	High-power AlGaInP three-ridge type laser diode array	<i>ibid.</i>	326-327	1988
H. van Houten, B. J. van Wees*, J. E. Mooij* (*Univ. of Technol., Delft), C. W. J. Beenakker, J. G. Williamson & C. T. Foxon	E,R	Coherent electron focussing in a two-dimensional electron gas	Europhys. Lett. 5	721-725	1988
P. Schobinger-Papamantellos (ETHZ, Zürich), K. H. J. Buschow, F. Maaroufi* & P. Toledano* (*Univ. Amiens)	E	Phenomenological theory of the temperature-composition phase-diagram in the magnetic system TbGe _{1-x} Si _x	Ferroelectrics 79	83-86	1988
R. Eppenga, S. Colak & M. F. H. Schuurmans	E	Effects of band mixing on the radiative properties of a quantum well	Fiber & Integrated Opt. 7	57-67	1988
R. Clasen	A	Preparation of glass and ceramics by sintering colloidal particles deposited from the gas phase	Glastech. Ber. 61	119-126	1988

- J. J. H. van den Biesen & T. Toyabe
(Central Res. Lab. Hitachi, Kokubunji, Tokyo) E Comparison of methods to calculate capacitances and cutoff frequencies from dc and ac simulations on bipolar devices IEEE Trans. CAD-7 855-861 1988
- D. J. Belton S The effect of post-mold curling upon the microstructure of epoxy molding compounds IEEE Trans. CHMT-10 358-363 1987
- S. A. Gee & C. G. M. van Kessel
(Philips Centre for Manufacturing Technol., Eindhoven) S Stiffness and yielding in the PLCC 'J' lead *ibid.* 379-390 1987
- V. Zieren, J. J. M. Ruigrok, M. J. Piena, S. B. Luitjens, C. W. M. P. Sillen & J. P. M. Verbunt E Efficiency improvement of one-sided probe heads for perpendicular recording on double-layer media IEEE Trans. MAG-23 2479-2481 1987
- F. J. A. den Broeder, D. Kuiper & H. J. G. Draaisma (Univ. of Technol., Eindhoven) E Effects of annealing and ion implantation on the magnetic properties of Pd/Co multilayers containing ultrathin Co *ibid.* 3696-3698 1987
- M. Moshfeghi N Sidelobe suppression in annular array and axicon imaging systems J. Acoust. Soc. Am. 83 2202-2209 1988
- H.-P. Stormberg, S. Murayama* & Y. Watanabe* (* Central Res. Lab. Hitachi, Kokubunji, Tokyo) A Angular distribution of x-ray radiation from optically thick z-pinch plasmas J. Appl. Phys. 62 4090-4095 1987
- M. Erman, C. Alibert (Equipe de Microoptoelectronique de Montpellier, Montpellier), J. B. Theeten, P. Frijlink & B. Catte L Continuous transition from multiple quantum-well regime to superlattice regime in GaAlAs/GaAs system as observed by spectroscopic ellipsometry with high lateral resolution J. Appl. Phys. 63 465-474 1988
- C. van Opdorp & M. J. Verkerk E Stress distribution in released vacuum-deposited aluminum films *ibid.* 1518-1525 1988
- E. K. Broadbent, J. M. Towner, P. H. Townsend (Stanford Univ., Stanford, CA) & H. A. Vander Plas (Xerox Palo Alto Res. Center, Palo Alto, CA) S Electromigration-induced short-circuit failure in aluminum underlaid with chemically vapor deposited tungsten *ibid.* 1917-1923 1988
- W. B. Zeper & A. M. J. Spruijt (Philips & Dupont Optical Company, Eindhoven) E Domain size measurements on GdTbFe-based thin-film structures for magneto-optical recording *ibid.* 2141-2146 1988
- D. J. Olego & H. Baumgart N Raman scattering characterization of the microscopic structure of semi-insulating polycrystalline Si thin films *ibid.* 2669-2673 1988
- K. H. J. Buschow E Structure and properties of some novel ternary Fe-rich rare-earth intermetallics *ibid.* 3130-3135 1988
- H. J. G. Draaisma*, F. J. A. den Broeder & W. J. M. de Jonge* (* Univ. of Technol., Eindhoven) E Perpendicular anisotropy in Pd/Co multilayers *ibid.* 3479-3481 1988
- O. B. Loopstra*, W. G. Sloof*, T. H. de Keijser*, E. J. Mittemeijer*, S. Radelaar* (* Univ. of Technol., Delft), A. E. T. Kuiper & R. A. M. Wolters E Composition, microstructure, and properties of crystalline molybdenum silicide thin films produced by annealing of amorphous Mo/Si multilayers *ibid.* 4960-4969 1988
- C. L. Kuo*, Y. H. Kao (State Univ. of New York at Stony Brook, Stony Brook, NY), E. Arnold, J. C. Bilello* (* California State Univ., Fullerton, CA) N Studies of lattice distortion and crystal curvature in arsenic-implanted silicon J. Appl. Phys. 64 1791-1794 1988
- R. M. Aarts & A. J. M. Kaizer E Simulation of loudspeaker crossover filters with a digital signal processor J. Audio Eng. Soc. 36 115-121 1988
- D. J. Verschuur*, A. J. M. Kaizer, W. F. Druyvesteyn & D. de Vries* (* Univ. of Technol., Delft) E Wigner representation of loudspeaker responses in a living room *ibid.* 203-212 1988
- S. M. Wolfgram & N. J. A. van Veen E Savitzky-Golay filtering in particle characterization J. Chemometrics 2 203-209 1988
- H. K. Kuiken N Fin methods of higher order J. Comput. Phys. 76 1-18 1988
- P. Capper, J. C. Brice, C. L. Jones*, W. G. Coates*, J. J. G. Gosney*, C. Ard* & I. Kenworthy* (*Mullard, Southampton) R Interfaces and flow regimes in ACRT grown Cd₂Hg_{1-x}Te J. Crystal Growth 89 171-176 1988

- | | | | | |
|--|--|---|-----------|------|
| M. Klee, W. Brand & J. W. C. de Vries
<i>E,A</i> | Superconducting films in the Y-Ba-Cu-O system made by thermal decomposition of metal carboxylates | J. Cryst. Growth 91 | 346-351 | 1988 |
| W. Busselt & R. Raue
<i>A</i> | Optimizing the optical properties of TV phosphor screens | J. Electrochem. Soc. 135 | 764-771 | 1988 |
| B. Greenberg & T. Marshall
<i>N</i> | The volume fraction of crystalline silicon in semi-insulating polycrystalline silicon (SIPOS) | <i>ibid.</i> | 2295-2299 | 1988 |
| G. M. Blom & J. M. Woodall (<i>IBM T. J. Watson Res. Center, Yorktown Heights, NY</i>)
<i>N</i> | Effect of iso-electronic dopants on the dislocation density of GaAs | J. Electron. Mater. 17 | 391-396 | 1988 |
| J. W. M. Jacobs
<i>E</i> | Note on a theory of three-dimensional electrochemical nucleation with diffusion-controlled growth | J. Electroanal. Chem. 247 | 135-144 | 1988 |
| K. A. Schouhamer Immink
<i>E</i> | Transmission of digital sound over channels with unforeseen bandwidth limitations | J. Inst. Electron. & Radio Eng. 58 | 49-56 | 1988 |
| C. J. M. Denissen, B. D. de Mooij & K. H. J. Buschow
<i>E</i> | Spin reorientation in Nd ₂ Fe ₁₄ C | J. Less-Common Met. 142 | 195-202 | 1988 |
| D. B. de Mooij & K. H. J. Buschow
<i>E</i> | Formation and magnetic properties of the compounds R ₂ Fe ₁₄ C | <i>ibid.</i> | 349-357 | 1988 |
| K. H. J. Buschow, D. B. de Mooij & C. J. M. Denissen
<i>E</i> | Magnetic properties and ⁵⁷ Fe Mössbauer effect in Nd ₂ Fe _{14-x} Mn _x C | <i>ibid.</i> | L13-L17 | 1988 |
| K. H. J. Buschow
<i>E</i> | Magnetovolume effects in ternary compounds of the type RFe _{12-x} T _x and R ₂ Fe ₁₄ C (R≡rare earth, T≡Ti, V, Cr, Mo, W, Si) | J. Less-Common Met. 144 | 65-69 | 1988 |
| B. H. Verbeek, C. H. Henry*, N. A. Olsson, K. J. Orlovsky*, R. F. Kazarinov* & B. H. Johnson* (<i>* AT&T Bell Labs, Murray Hill, NJ</i>)
<i>E</i> | Integrated four-channel Mach-Zehnder multi/demultiplexer fabricated with phosphorous doped SiO ₂ waveguides on Si | J. Lightwave Technol. LT-6 | 1011-1015 | 1988 |
| G. J. van der Kolk
<i>E</i> | Crystallization temperatures of 5d-5d alloys | J. Mater. Res. 3 | 209-211 | 1988 |
| F. J. C. M. Toolenaar & M. T. J. Verhees
<i>E</i> | Reactive sintering of zinc ferrite | J. Mater. Sci. 23 | 856-861 | 1988 |
| P. K. Bachmann, D. U. Wiechert & T. P. M. Meeuwssen
<i>E,A</i> | Thermal expansion coefficients of doped and undoped silica prepared by means of PCVD | <i>ibid.</i> | 2584-2588 | 1988 |
| R. Clasen
<i>A</i> | Preparation of high-purity silica glass tubes by centrifugal casting of colloidal gels | J. Mater. Sci. Lett. 7 | 477-478 | 1988 |
| K. J. B. M. Nieuwesteeg
<i>E</i> | The influence of realistic velocity-dependent cross sections on the correlation between Doppler and collisional broadening of the Na D ₁ spectral line | J. Phys. B 21 | 1353-1365 | 1988 |
| J. E. Griffith*, J. A. Kubby*, P. E. Wierenga, R. S. Becker* & J. S. Vickers* (<i>* AT&T Bell Labs, Murray Hill, NJ</i>)
<i>E</i> | Tunneling microscopy of steps on vicinal Ge(001) and Si(001) surfaces | J. Vac. Sci. & Technol. A 6 | 493-496 | 1988 |
| N. Vodjdani & P. Parrens (<i>Nextral, Meylan</i>)
<i>L</i> | Reactive ion etching of GaAs with high aspect ratios with Cl ₂ -CH ₄ -H ₂ -Ar mixtures | J. Vac. Sci. & Technol. B 5 | 1591-1598 | 1987 |
| M. F. C. Willemsen, A. E. T. Kuiper, A. H. Reader, R. Hokke & J. C. Barbour
<i>E</i> | <i>In situ</i> investigation of TiN formation on top of TiSi ₂ | J. Vac. Sci. & Technol. B 6 | 53-61 | 1988 |
| H. J. Cornelissen, D. A. Cammack & R. J. Dalby
<i>N</i> | Reflection high-energy electron diffraction observations during growth of ZnS ₂ Se _{1-x} (0≤x≤1) by molecular-beam epitaxy | <i>ibid.</i> | 769-772 | 1988 |
| D. J. Olego
<i>N</i> | Effects of II-VI epitaxy on III-V surfaces: a study of ZnSe on GaAs | <i>ibid.</i> | 1193-1197 | 1988 |
| M. Erman, M. Renaud & S. Gourrier
<i>L</i> | Spatially resolved ellipsometry for semiconductor process control: application to GaInAs MIS structures | Jap. J. Appl. Phys. 26 | 1891-1897 | 1987 |
| J. Terken* & S. G. Nooteboom* (<i>* Inst. for Perception Res., Eindhoven</i>) | Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information | Lang. & Cognitive Processes 2 | 145-163 | 1987 |

- P. Capper, P. A. C. Whiffin, B. C. Easton, C. D. Maxey & I. Kenworthy (Mullard, Southampton) *R* Group V acceptor doping of $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ layers grown by metal-organic vapour phase epitaxy *Mater. Lett.* **6** 365-368 1988
- B. Dam, T. S. Baller, G. N. A. van Veen, H. A. M. van Hal, J. W. C. de Vries & G. M. Stollman *E* Interface effects of laser ablated and sputtered high T_c superconducting thin films *Mater. Res. Soc. Symp. Proc.* **99** 303-306 1988
- R. J. Dalby & J. Petruzzello *N* Asymmetric cracking in $\text{ZnSe}/\text{ZnS}_x\text{Se}_{1-x}$ superlattices grown by molecular beam epitaxy *Mater. Res. Soc. Symp. Proc.* **102** 131-136 1988
- B. A. Khan & R. Pandya *N* Effects of hydrogenation on the off-state of polysilicon thin film transistors *Mater. Res. Soc. Symp. Proc.* **106** 353-358 1988
- T. Marshall, E. Arnold & B. Khan *N* Electrical characterization of Si-SiO₂ interfaces in thin-film SOI structures *Mater. Res. Soc. Symp. Proc.* **107** 353-357 1988
- H. Greiner *A* Computing complex chemical equilibria by generalized linear programming *Math. & Comput. Modelling* **10** 529-550 1988
- W. Hillen, U. Schiebel & T. Zaengel *A* Imaging performance of a digital storage phosphor system *Med. Phys.* **14** 744-751 1987
- F. J. J. Blommaert* & H. Timmers* (*Inst. for Perception Res., Eindhoven) Letter recognition at low contrast levels: effects of letter size *Perception* **16** 421-432 1987
- R. van de Plassche & P. Baltus *E,S* The design of an 8-bit folding analog-to-digital converter *Philips J. Res.* **42** 482-509 1987
- F. M. J. de Bont, E. H. L. Aarts, P. Meehan* C. G. O'Brien* (*Trinity College, Dublin) *E* Placement of shapeable blocks *Philips J. Res.* **43** 1-22 1988
- A. J. E. M. Janssen *E* The Zak transform: a signal transform for sampled time-continuous signals *ibid.* 23-69 1988
- D. B. de Mooij & K. H. J. Buschow *E* Note on the structure and composition of the B-rich ternary phase in the Nd-Fe-B system *ibid.* 70-74 1988
- G. F. M. Beenker & J. H. van Lint *E* Optimal generalized Petersen graphs *ibid.* 129-136 1988
- P. J. Severin & W. H. Bardeel *E* A simple multimode fibre interferometric sensor for pressure-related measurements *ibid.* 137-151 1988
- H. E. Wanders (Philips Centre for Quantitative Methods, Eindhoven) *E* On the significance of Golomb's randomness postulates in cryptography *ibid.* 185-222 1988
- C. W. J. Beenakker *E* Numerical simulation of a coarsening two-dimensional network *Phys. Rev. A* **37** 1697-1702 1988
- I. Heynderickx & H. de Raedt (Univ. Antwerp & Univ. Amsterdam) *E* Calculation of the director configuration of nematic liquid crystals by the simulated-anneal method *ibid.* 1725-1730 1988
- A. D. C. Grassie*, K. M. Hutchings*, M. Lakrimi* (*Univ. of Sussex, Brighton), C. T. Foxon & J. J. Harris *R* Shubnikov-de Haas effect in submicron-width GaAs/Ga_{1-x}Al_xAs heterojunction wires *Phys. Rev. B* **36** 4551-4554 1987
- H. van Houten, J. G. Williamson, M. E. I. Broekaat, C. T. Foxon & J. J. Harris *E,R* Magnetoresistance in a GaAs-Al_xGa_{1-x}As heterostructure with double subband occupancy *Phys. Rev. B* **37** 2756-2758 1988
- G. E. W. Bauer & T. Ando (Univ. Tokyo) *E* Theory of magnetoexcitons in quantum wells *ibid.* 3130-3133 1988
- R. Fletcher (Queen's Univ., Kingston, Ontario), M. D'Iorio*, A. S. Sachrajda*, R. Stoner* (*Nat. Res. Council, Ottawa, Ontario), C. T. Foxon & J. J. Harris *R* Evidence of phonon drag in the thermopower of a GaAs-Ga_{0.68}Al_{0.32}As heterojunction *ibid.* 3137-3140 1988
- C. W. J. Beenakker & H. van Houten *E* Flux-cancellation effect on narrow-channel magnetoresistance fluctuations *ibid.* 6544-6546 1988
- R. M. Tromp*, E. J. van Loenen, J. E. Demuth* & N. D. Lang* (*IBM T. J. Watson Res. Center, Yorktown Heights, NJ) *E* Tip electronic structure in scanning tunneling microscopy *ibid.* 9042-9045 1988
- K. Shahzad, D. J. Olego & C. G. van de Walle (IBM T. J. Watson Res. Center, Yorktown Heights, NY) *N* Optical characterization and band offsets in ZnSe-ZnS_xSe_{1-x} strained-layer superlattices *Phys. Rev. B* **38** 1417-1426 1988

M. H. Frommer	N	X-ray photoelectron investigation of the valence-band electronic structure of superconducting Y-Ba-Cu oxide	Phys. Rev. B 38	2444-2449	1988
R. Eppenga & M. F. H. Schuurmans	E	Thin [001] and [110] GaAs/AlAs superlattices: distinction between direct and indirect semiconductors	<i>ibid.</i>	3541-3544	1988
J. Khurgin	N	Second-order nonlinear effects in asymmetric quantum-well structures	<i>ibid.</i>	4056-4066	1988
D. J. Olego, K. Shahzad, D. A. Cammack & H. Cornelissen	N	Optical vibrational modes of ZnSe-ZnS _x Se _{1-x} strained-layer superlattices	<i>ibid.</i>	5554-5564	1988
K. Shahzad	N	Excitonic transitions in ZnSe epilayers grown on GaAs	<i>ibid.</i>	8309-8312	1988
H. J. Kappen	E	Reply to "Comment on 'Calculating the weak scale in supergravity models'"	Phys. Rev. D 38	721-722	1988
M. D. Pashley, K. W. Haberern, W. Friday, J. M. Woodall* & P. D. Kirchner* (* IBM T. J. Watson Res. Center, Yorktown Heights, NJ)	N	Structure of GaAs(001) (2x4)-c (2x8) determined by scanning tunneling microscopy	Phys. Rev. Lett. 60	2176-2179	1988
J. W. Severin, G. de With & H. A. M. van Hal	E	Low-temperature synthesis route for YBa ₂ Cu ₃ O _x powder	Physica C 152	159-160	1988
G. N. A. van Veen, T. S. Baller, J. W. C. de Vries & G. M. Stollmann	E	Degradation of oriented superconducting thin films	<i>ibid.</i>	267-271	1988
D. J. Belton & M. Molter	S	Concurrent enthalpy recovery and curing phenomena in epoxy encapsulating materials for semiconductor applications	Polym. Eng. & Sci. 28	189-199	1988
J. Lub, F. C. B. M. van Vroonhoven, D. van Leyen* & A. Benninghoven* (*Univ. Münster)	E	Static secondary ion mass spectrometry analysis of polycarbonate surfaces. Effect of structure and of surface modification on the spectra	Polymer 29	998-1003	1988
W. S. Newman & N. Hogan (Massachusetts Inst. of Technol., Cambridge, MA)	N	High speed robot control and obstacle avoidance using dynamic potential functions	Proc. IEEE Int. Conf. on Robotics and automation, Raleigh, NC, 1987	14-24	1987
M. Rocchi	L	A critical review of GaAs digital IC's design and performances for computer applications	Proc. IEEE Int. Conf. on Computer design: VLSI in computers, Port Chester, NY, 1986	131-136	1986
E. K. Broadbent, D. K. Sadana, A. E. Morgan, J. M. Flanner & R. C. Ellwanger	E,S	Some recent observations on tunnel defect formation during high temperature post-deposition anneal of CVD W on Si	Proc. Workshop on Tungsten and other refractory metals for VLSI applications, Yorktown Heights, NY, 1987	111-113	1988
E. K. Broadbent, J. M. Flanner, W. G. M. van den Hoek & I.-W. H. Connick	S	High density, high reliability tungsten interconnection by filled interconnect groove (FIG) metallization	<i>ibid.</i>	191-199	1988
R. Vijaykumar*, S. Venkataraman*, G. Dakin* (*Univ. of Massachusetts, Amherst, MA) & D. M. Lyons	N	A task grammar approach to the structure and analysis of robot programs	Proc. IEEE Workshop on Languages for automation, Vienna 1987	39-42	1987
K. Schroeder Trovato & W. Schreiner	N	LABICS: a language for a distributed, hierarchical, and dynamically reconfigurable control system	<i>ibid.</i>	43-47	1987
D. M. Lyons	N	A novel approach to high-level robot programming	<i>ibid.</i>	48-51	1987
D. Washington	R	Color display using the channel multiplier CRT	Proc. SID 29	23-31	1988
S. Herman	N	An introduction to sifting theory	Proc. SPIE 767	332-337	1987
P. Blood, E. D. Fletcher & C. T. Foxon	R	All-binary short period GaAs-AlAs superlattice quantum well lasers grown by molecular beam epitaxy	Proc. SPIE 861	61-66	1988
W. Hillen, U. Schiebel & T. Zaengel	A	A selenium-based detector system for digital slot-radiography	Proc. SPIE 914	253-261	1988
S. Herman	N	Feature-size dependent selective edge enhancement of x-ray images	<i>ibid.</i>	654-659	1988

R. B. Beall, J. B. Clegg & J. J. Harris R	Migration of Si in δ -doped GaAs	Semicond. Sci. Technol. 3	612-615	1988
D. E. Lacklison, J. J. Harris, C. T. Foxon, J. Hewett, D. Hilton & C. Roberts R	A comparison of photoconduction effects in (Al,Ga)As and GaAs/(Al,Ga)As heterostructures	<i>ibid.</i>	633-640	1988
J. J. Harris, J. M. Lagemaat, S. J. Battersby, C. M. Hellon, C. T. Foxon & D. E. Lacklison E,R	Sub-band populations and the spatial distribution of electrons in GaAs/(Al,Ga)As modulation-doped quantum wells	<i>ibid.</i>	773-780	1988
F. Grainger & J. A. Roberts R	Impurity survey analysis of semiconductor layers by laser scan mass spectrometry	<i>ibid.</i>	802-807	1988
P. W. J. M. Boumans E	Needs for fundamental atomic reference data and data distribution: applied analytical spectroscopy — response to a questionnaire	Spectrochim. Acta 43B	5-13	1988
P. W. J. M. Boumans & J. J. A. M. Vrakking E	Line selection in atomic emission spectrometry using physically resolved spectra	<i>ibid.</i>	69-74	1988
P. W. J. M. Boumans E	Needs for fundamental atomic reference data and data distribution — a collage of discussions	<i>ibid.</i>	103-127	1988
S. Colak N	An effective potential approach for extended density of states in heterostructures	Superlattices & Microstruct. 4	571-576	1988
T. Beelen (<i>Philips PICOS Glass, Eindhoven</i>) & P. van Dooren B	A numerical method for deadbeat control of generalized state-space systems	Syst. & Control Lett. 10	225-233	1988
C. Ronse B	Codage en liste d'arbres quaternaires	Tech. & Sci. Inf. 7	235-245	1988
B. A. Joyce, P. J. Dobson & P. K. Larsen E,R	Molecular beam epitaxy of III-V compounds. Aspects of growth kinetics and dynamics	The chemical physics of solid surfaces and heterogeneous catalysis, D. A. King & D. P. Woodruff (eds), Vol. 5, Surface properties of electronic materials, Elsevier Science, Amsterdam	271-307	1988
R. Eppenga E	The electronic and optical properties of GaAs/AlGaAs quantum wells	Thesis, Delft	1-121	1988
H. van Wijngaarden (<i>Philips Centre for Manufacturing Technol., Eindhoven</i>) E	Constitutive equations for metals with an application to the extrusion of lead	Thesis, Eindhoven	1-116	1988
H. J. G. Draaisma E	Magnetic-multilayers	<i>ibid.</i>	1-162	1988
I. J. M. M. Raaijmakers E	Fundamental aspects of reactions in titanium-silicon thin films for integrated circuits	<i>ibid.</i>	1-141	1988
R. N. J. Veldhuis E	Adaptive restoration of unknown samples in discrete-time signal and digital images	Thesis, Nijmegen	1-159	1988
J. W. M. Jacobs E	Laser-initiated metal deposition on semiconductors from aqueous solutions	Thesis, Utrecht	1-221	1988
R. Woltjer E	The quantized Hall effect due to inhomogeneities	<i>ibid.</i>	1-119	1988
P. F. Fewster R	Probing semiconductor MQW structures by X-ray diffraction	Thin film growth techniques for low-dimensional structures, R. F. C. Farrow <i>et al.</i> (eds), Plenum, New York	417-440	1987
J. P. Gowers R	The TEM characterisation of low-dimensional structures in epitaxial semiconductor thin film	<i>ibid.</i>	471-486	1987
R. C. Ellwanger & J. M. Towner S	The deposition and film properties of reactively sputtered titanium nitride	Thin Solid Films 161	289-304	1988
M. Moshfeghi & R. C. Waag (<i>Univ. of Rochester, Rochester, NY</i>) N	<i>In vivo</i> and <i>in vitro</i> ultrasound beam distortion measurements of a large aperture and a conventional aperture focussed transducer	Ultrasound Med. & Biol. 14	415-428	1988



L. Vriens, J. A. Clarke and J. H. M. Spruit, Interference filters in projection television tubes,
Philips Tech. Rev. 44, No. 7, 201-210, March 1989.

A multilayer interference filter — a stack of layers of alternately high and low refractive indices — alters the distribution of the light emitted by the phosphor on the faceplate of a television tube in such a way that more light is emitted in the forward direction: When these filters are used in projection television tubes, more light falls within the acceptance angle of the lens system, so that more light contributes to the brightness on the projection screen. The brightness everywhere on the screen can be improved by using curved faceplates in the tubes. The chromaticity, resolution and contrast in the projected picture are also improved when such filters are used.

J. L. van Meerbergen and H. De Man, A true silicon compiler for the design of complex ICs for digital signal processing,
Philips Tech. Rev. 44, No. 7, 218-230, March 1989.

PIRAMID is a sophisticated software package ('silicon compiler') for the highly automated design of complex integrated circuits (ICs) for digital signal processing in areas such as speech processing, telecommunications, audio and lower-end image processing. A general target architecture for IC design work has been defined, with a data path consisting of a number of execution units and a controller which controls the data path via a microcode. PIRAMID consists of three environments: architectural synthesis (based on CATHEDRAL II computer programs), floorplanning and module generation. The silicon compilation starts from a system description in the high-level computer language Silage and a library of elementary units ('leaf cells') that contain all the relevant design information about the IC technology to be used; it finishes with a fully detailed IC-layout database. The module-generation environment makes use of the leaf cells to design the execution units and the controller, with the parameter values (such as wordlengths, memory capacities, number of registers in a register file) specified by the other two environments. The designer can 'steer' the compilation process by giving specific hints ('pragmas'). The main features of PIRAMID are illustrated in this article by tracing through the different stages in the IC design for a 64-tap adaptive digital filter.

J. F. Dijkstra, Analysis of the injection-moulding process,
Philips Tech. Rev. 44, No. 7, 212-217, March 1989.

This article is based on a talk given at a colloquium held in May 1988 at Philips Research Laboratories. A brief review of a number of plastic products is followed by a short account of the injection-moulding process. This process is analysed as it would be seen by several different specialists: the materials scientist (amorphous plastics, monomers, mixtures, etc.), the product designer (mechanical strength, temperature range, sensitivity to moisture, etc.), the mould designer (dimensions of the cavity in connection with shrinkage, runner systems, separating the mould, etc.) and the rheologist (coefficient of expansion, viscosity, thermal conduction, etc.). The relationship between shaping and the properties of the product is also examined. The article ends with a few words on the help that software can offer in predicting the fill pattern, packing and cooling as well as errors in shape and dimensioning.

Contents

- | | |
|---|-----|
| 75 years of Philips Research
<i>editorial</i> | 237 |
| 75 years of research: from lamps to integrated circuits
P. Kramer | 239 |
| Manufacturing optical fibres by the PCVD process
P. Geittner and H. Lydtin
<i>Chemical deposition from the gas phase can be used
for manufacturing high-quality optical fibre
for telecommunications</i> | 241 |
| Phase-change optical recording
D. J. Gravesteijn, C. J. van der Poel, P. M. L. O. Scholte and
C. M. J. van Uijen
<i>Materials that change to another phase when heated by a laser beam
are of considerable interest for the recording and erasure
of information on an optical disc</i> | 250 |
| Magnetic fields in medical diagnostics: MR and SQUID
O. Dössel, M. H. Kuhn and H. Weiss
<i>Magnetic fields of up to four teslas present a considerable challenge in MR;
the challenge in SQUID is in the measurement of fields
only 10^{-14} of this value</i> | 259 |

The application of semiconductor superlattices to short-wavelength lasers	268
P. Blood, C. T. Foxon and E. D. Fletcher	
<i>Superstructures consisting of alternate ultra-thin layers of GaAs and AlAs can be used to advantage in semiconductor lasers for operation at wavelengths below 880 nm</i>	
Predicting the properties of materials: dream or reality?	276
M. F. H. Schuurmans, R. Coehoorn, R. Eppenga and P. J. Kelly	
<i>In calculating the properties of materials with the density functional theory the only input required is the atomic number</i>	
A sound basis for the generation of explanations in Expert Systems	287
A. Bruffaerts, E. Henin and A. Pirotte	
<i>Users will have even more confidence in an Expert System that can explain its answers</i>	
HS3: an advanced bipolar-IC technology	296
G. Conner and R. H. Lane	
<i>HS3 combines versatility with the ability to provide circuits with excellent high-frequency behaviour and direct-current characteristic parameters</i>	
Research on monolithic GaAs MESFET circuits at LEP	302
M. Rocchi	
<i>The potentialities of ICs based on GaAs field-effect transistors have vastly increased in the last fifteen years</i>	
Power integrated circuits	310
M. Amato, G. Bruning, S. Mukherjee and I. T. Wacyk	
<i>The ability to combine power circuits with other circuits on a single chip will revolutionize some products</i>	
Scientific publications	321

75 years of Philips Research

The 75th birthday of Philips Research is clearly an appropriate occasion for a special issue of our journal. An anniversary like this prompts a little nostalgia as we look back, and this is understandable, since research has changed so greatly in those seventy-five years. At first the only research was on tungsten incandescent lamps, but as the years went by many other fields had to be tackled as well: gas discharges, medical systems, electric motors, radio, television, magnetic and optical recording, telephony, integrated circuits, and more besides. All this growth has had important consequences for the numbers of people and their working environment.

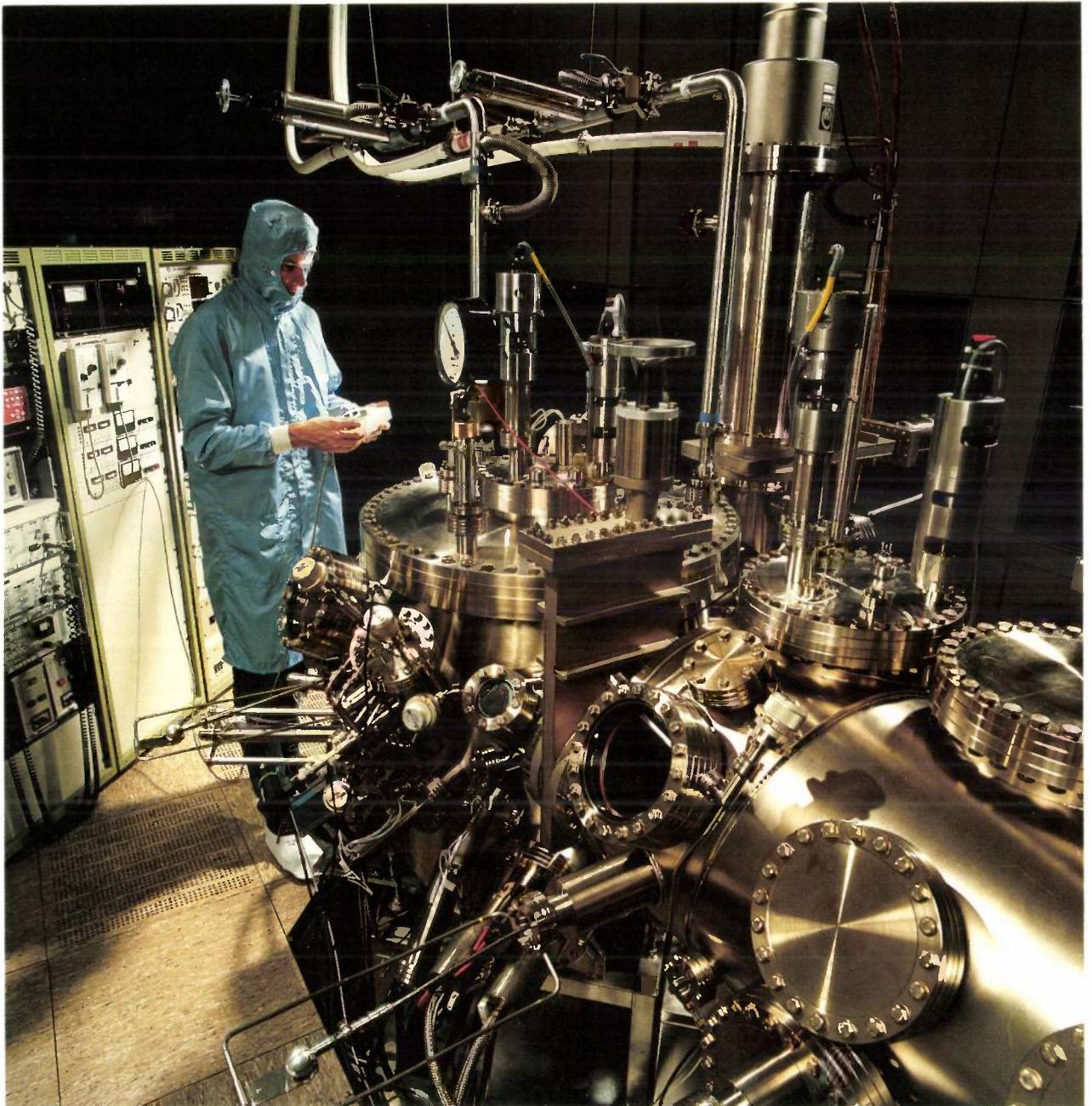
In the early years there were only a few dozen staff, and the 'Philips Natuurkundig Laboratorium' (literally the 'Philips Physics Laboratory') was still part of the Eindhoven lamp works. In 1923 the Laboratory moved to a separate building, which was considerably extended later. In 1939 there were nearly 500 people working in the laboratory, by now more widely known in the English-speaking world as 'Philips Research Laboratories'. After World War II research grew rapidly and new quarters were obviously necessary. The outcome was the construction of the present buildings, still known within the Company as 'the Nat. Lab.', on the outskirts of Eindhoven. The new Laboratories were officially opened in 1963, and many new buildings have been added since. Laboratories have also been established in West Germany (in Aachen and Hamburg), the United Kingdom (at Redhill), France (at Limeil-Brévannes), Belgium (in Brussels) and the United States (at Briarcliff and Sunnyvale). Now Philips Research employ some 4200 people, about half of them at the Research Laboratories in Eindhoven.

From its first issue in 1936 our journal has always given a detailed account of the results of Philips Research. In this issue we shall try to give a representative picture of the research at this moment. A characteristic subject has been selected from each of the Laboratories, with two articles representing the Eindhoven Laboratories. We think the result shows the strength of Philips Research and gives a good picture of its wide variety.

A few personal notes by Dr P. Kramer, until recently General Director of Philips Research, are followed by two articles that pay particular attention to the research on materials for applications that have been extensively developed in recent years: optical fibres for communications and new materials for optical recording. The next article is about recent advances in medical diagnostics by magnetic resonance (MR). Then we have three subjects that are closer to fundamental research: the study of special semiconductor structures, the prediction of properties of new materials and the development of computer programs for artificial intelligence. Finally, three articles describe various kinds of integrated circuits. Here again there is great variety, both in the basic materials (silicon and GaAs) and in the types of circuits (bipolar ICs, field-effect-transistor ICs and power ICs).

Our special issue is a reminder that Philips Research, which started as a small department for investigating incandescent lamps, has grown into a large but versatile organization. This research is vital to the continuity of Philips as an international science-based company.

The editors



Molecular beam epitaxy is a very refined form of ultra-high vacuum evaporation. The technique can be used for applying very thin layers of material to a substrate; the layers can have thicknesses of the order of nanometres and can include abrupt transitions in composition. A multilayer structure of Si and Ge with sharp doping profiles, deposited on a substrate, can be used to make better (and indeed new) semiconductor components. The photograph shows a silicon-MBE equipment in a dust-free room at Philips Research Laboratories in Eindhoven. The actual evaporation takes place in the large vertical ultra-high-vacuum vessel beside the scientist. A spectroscopic analysis of the substrate or the evaporated layer can be made before or after the evaporation, in the horizontal cylinder partly visible at the lower right. The large racks in the background contain measurement and control equipment.

75 years of research: from lamps to integrated circuits

P. Kramer

January 2nd 1914 can be taken as the starting point for research at Philips. This was when Gilles Holst^[1] joined the 'Naamlooze Vennootschap Philips' Metaal Gloeilampenfabriek'^[2] to do research on incandescent lamps and gas discharges. If I now look back over those 75 years of research, I can see two important evolutionary developments: Philips has gone from local lamp works to international electronics company, and research on discrete components has given way to research on complex electronic systems.

Work in lamp technology led to a good understanding of materials, vacuum technology and glass processing. In the period between the two World Wars this in turn led to research on X-ray tubes and thermionic valves. This was followed eventually by research on television picture tubes, image-intensifiers, and so on. The birth of radio called for research on electronic circuits. At first the circuits were simple, with only a few components.

Soon after World War II the invention of the transistor unleashed a real revolution. Research on transistors required a deep understanding of materials, especially semiconductors. In circuit development there were many more components to choose from than there had been in the valve era. As well as circuits for radio and television, new circuits — much larger ones — were now designed for computers. Electronic circuits could be much more complex, since a transistor was small, cheap and reliable, and it dissipated much less power than a valve.

The real large-circuit breakthrough came in about 1960, when it became possible to integrate planar transistors and other components on silicon, at first as analog circuits with bipolar transistors. At Philips, the first application was an amplifier circuit for a consumer product, a hearing aid. Later on, digital integrated circuits with MOS transistors (MOS stands for Metal-Oxide Semiconductor) were widely applied in

computers. The number of transistors per circuit meanwhile increased steadily; we had in turn LSI, VLSI and ULSI (Large-Scale, Very-Large-Scale and Ultra-Large-Scale Integration).

With the almost exponential increase in the number of functions on an IC — in digital memory circuits, for example, the numbers of cells quadruple about every three years — the number of possible applications also increases strongly. Digital ICs are not only used in the obvious applications such as computing, but in electronic filters and in other signal-processing operations as well. Integrated circuits are also used for digital signal processing in Compact Disc players and television sets. So we have digital ICs in consumer products, and the wheel of applications has swung full circle. The potential of integrated circuits is now vast. Not only is the frequency of the signals that can be processed much higher; the currents and voltages that can be handled are higher too. Complete power supplies can now be integrated on silicon, for example. When these are combined with low-voltage circuits, we have 'smart power' chips, which can be used in applications such as power supplies for discharge lamps and electric motors.

The 'driving force' behind the growth in the number of applications for ICs is the continuous rise in the number of functions for the same price, both per unit area of silicon. And this growth is not limited by problems of energy or environment. The problems that do exist are more likely to be found in putting the IC technology into practice. As the details of the ICs become smaller, the number of mask steps increases, as do the accuracy and control required at each step. Nor is it at all easy to design circuits containing millions of transistors. A coherent package of powerful computer programs is required, with good interfacing

[1] H. B. G. Casimir, Gilles Holst, pioneer of industrial research in the Netherlands, Philips Tech. Rev. 40, 121-128, 1982.

[2] This means something like 'Philips Metal Incandescent Lamp Works Ltd'.

between the various levels of the design hierarchy. The creation of software tools for modelling, verification and simulation of integrated circuits calls for considerable effort from many people.

The transistors and other components of an electronic circuit cannot be integrated on silicon without certain other kinds of integration too:

- integration of the specialists who are collectively responsible for the design, technology and application of the circuit in a system,
- integration of the organization, from research laboratories to sales departments,
- integration of the carefully planned activities, which have to be carried out in teams and projects.

This integration of people and their activities means that everybody must agree at a very early stage on the design rules and protocols, and on the commercial strategies as well. An idea originating from a research laboratory must be evaluated for ease of production and possible consequences related to investment. Brilliant researchers may therefore encounter obstacles to the adoption of their ideas.

So integration here means that innovations in the IC industry must tie up with developments in the past. The picture of a steamroller comes immediately to mind here: progress cannot be held back and its direction and speed are predictable. If we think of this picture for a moment, we may wonder if there is still any place for research on integrated circuits:

- *on the steamroller*, working on the technology,
- *in front of it*, working on the applications.

This special issue of Philips Technical Review should make it plain that there is still plenty of room at Philips for research on integrated circuits: almost half of the articles describe research on the design, fabrication and application of ICs. This strong position for research in these fields makes sense, because of the creation of new ideas from all the new opportunities that microelectronics can offer. The digital signal processing I mentioned earlier has created new methods for image-processing techniques. In medical applications it has brought new fields into being — just think of X-ray computed tomography and diagnostics by proton spin resonance. Digital-processing techniques are also useful in consumer electronics, too: high-definition television, for example, would be impossible without these techniques.

Other activities that would be impossible now or in the future without these techniques include:

- picture recognition,
- automated production systems,
- database management,

- machine translation,
- artificial intelligence,
- parallel computer systems, and
- neural networks.

All these activities in turn provide stimuli for research '*in front of the steamroller*'.

Research fields '*on the steamroller*' are the mastery of new and existing manufacturing processes, and the development of software tools — 'silicon compilers' — for the design of integrated circuits. Modelling adaptive manufacturing processes for the new generations of ICs is a separate subject, 'manufacturing science'. Other research is necessary to overcome the obstacles that bar our way towards ever-smaller details on the silicon surface. And we also have to think of small details *in* the silicon, since there is research on three-dimensional structures too. The great question is, when are we going to reach the physical limits.

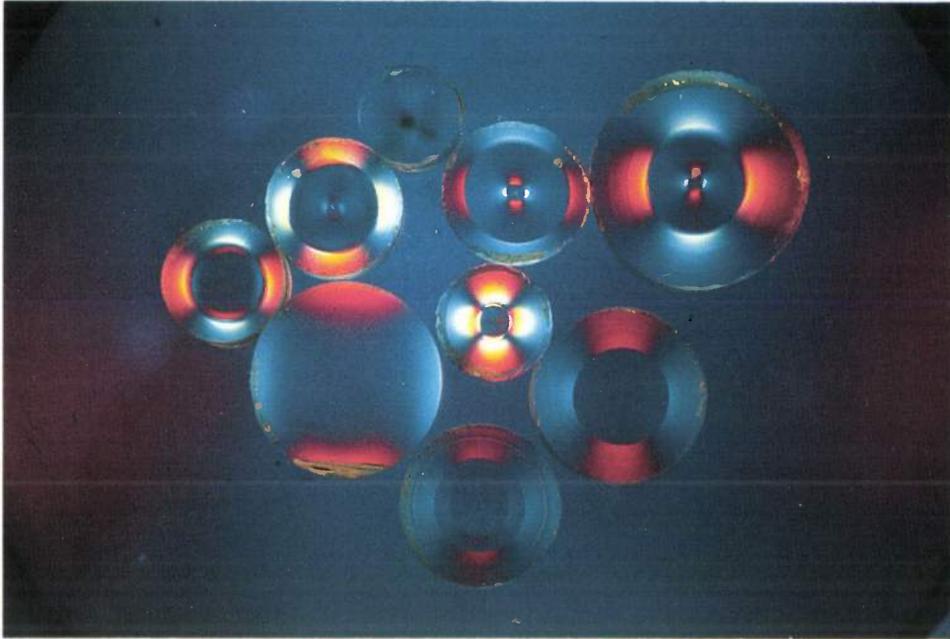
Another question, and a fundamental one, is whether there is an alternative to silicon. We know gallium arsenide, which is already used for transistors, 'ordinary' semiconductor lasers and pulsed semiconductor lasers. Then there are the developments in two-dimensional electron gases with very high electron mobilities and in ballistic electron effects. Finally, there are speculative ideas on the application of molecular or biological 'systems' for switching or memory functions.

The quest for fast switching effects and smaller new memory elements calls for individual effort from keen researchers in many fields. Further development is usually a group effort, where the most important steps are in the design of manufacturing processes, large circuits and large systems. Besides the exhilarating individual quest for a breakthrough in discrete components, which was characteristic of the first phase of research at Philips, there is plenty of scope now for team spirit in providing new world standards and new architectures. The pattern of research work is different now. Today's researcher is usually part of a large team, cooperating with colleagues from various disciplines. The research results must be transferred to development groups in good time and the quality of the research should make redesign unnecessary.

Life in the laboratories of a world-wide electronics company has changed enormously in 75 years. The complexity of large research programmes has led to a close cooperation between researchers, often with firm leadership. The researchers are as likely to draw inspiration from a common objective as from an individual goal. Pioneers can cooperate too!

Manufacturing optical fibres by the PCVD process

P. Geittner and H. Lydtin



Copper wires or coaxial cables for communications are increasingly being replaced by optical fibres made of glass. The main reasons for this are the increased signal frequency and the lower attenuation. Ten years ago an attenuation of 1 dB/km might have seemed impressive; today an attenuation of 0.2 dB/km is attainable. The advantage of the very low attenuation is that fewer repeaters are required in transmission links, so that the costs per kilometre of transmission path diminish. Other advantages of optical fibres are their insensitivity to interference from electromagnetic fields and the ready availability of the basic raw material — silica. Optical fibres with very low attenuation can only be made from extremely pure materials, however, and doping with special additives must be controlled very accurately. In the PCVD process (Plasma-activated Chemical Vapour Deposition), the result of research at Philips Forschungslaboratorium Aachen, these very difficult conditions can be met.

Introduction

Optical methods for transmitting messages have been known for many thousands of years. In all such methods there is always a transmitter, a transmission path and a receiver. The transmitter can be thought of as the person sending smoke or light signals, the receiver as the person who observes these signals and the transmission path as the atmosphere between the transmitter and the receiver. Two fairly advanced forms of

optical communication were Claude Chappe's visual (or semaphore) telegraph system^[1], which was set up in France in the mid-19th century, and A. G. Bell's 'photophone'^[2], see *fig. 1*. In spite of the enormous velocity of light the transmission rate was always limited until fairly recently by the inertia of the transmit-

Dr P. Geittner and Dr H. Lydtin are with Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany.

^[1] G. R. M. Garratt, The early history of telegraphy, Philips Tech. Rev. 26, 268-284, 1965.

^[2] C. Mackenzie, Alexander Graham Bell, Grosset and Dunlap, New York 1928.

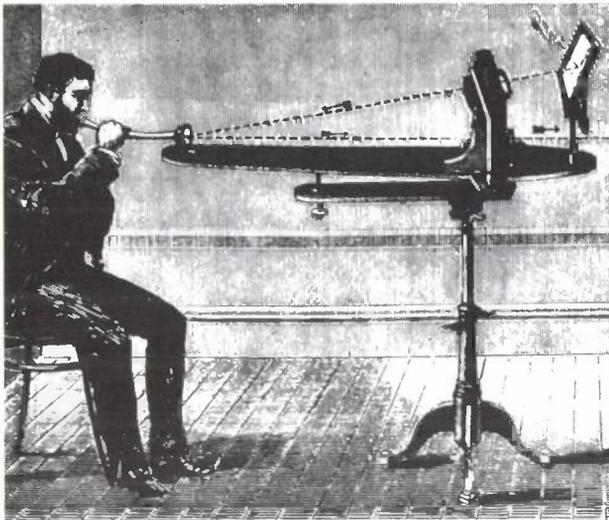


Fig. 1. The transmitter of the 'photophone', one of Alexander Graham Bell's inventions^[2]. The intensity of a beam of sunlight is modulated by reflecting it from a mirror attached to the end of a speaking tube. The beam falls on the receiver, a detector in the form of a layer of selenium, whose resistance is a function of the intensity of the incident light. The variation in resistance is converted into sound with the aid of a telephone. Bell's photophone could operate over a distance of several hundred metres. The first message it transmitted was the voice of his assistant Sumner Tainter, saying 'Mr. Bell, Mr. Bell, if you hear what I say, come to the window and wave your hat!'

ter and the receiver and by undesirable effects along the transmission path, such as fog, rain or other kinds of bad weather.

Attempts were made to solve the problem of these atmospheric effects by using enclosed systems with mirrors or lenses or light-guiding tubes, but there was no real answer until light was successfully propagated over long distances in optical fibres^[3]. In the sixties semiconductor lasers and photodiodes that could be used as transmitters and receivers became available, see fig. 2. The breakthrough came in 1970, when it was shown that the attenuation in an optical fibre could be less than 20 dB/km^[5].

Since optical transmission had such great advantages compared with systems that used copper wire, extensive research and development activities started up in most of the industrialized countries. The results of this can be seen today from the continuing growth of optical transmission. For transmitting information over long distances optical systems are already cheaper than coaxial-cable systems. This will probably not be true for local transmission links until the nineties.

The economic advantages come from the reduced number of repeaters compared with coaxial systems. Fewer repeaters are required because the attenuation of optical fibres has fallen considerably in recent years. The total amount of optical fibre manufactured, cabled and installed for communications is now 2 to 3 million kilometres a year. In spite of intensive

research on other processes there are at the moment only four methods mature enough for production. These are all based on the CVD process (Chemical Vapour Deposition), in which material is deposited from a reactive gas mixture. Philips manufacture optical fibres by a modified version of this method, the PCVD process (Plasma-activated Chemical Vapour Deposition).



Fig. 2. Schematic representation of modern optical transmission. *S* transmitter, a laser or LED (Light-Emitting Diode). *T* transmission path via an optical fibre. *R* receiver, an 'ordinary' semiconductor photodiode or an avalanche photodiode^[4].

In this article we first look at some of the background physics encountered in the transmission of information on optical fibres. Then we consider how this affects the requirements for the manufacturing process. Next we look at the actual Philips PCVD process and the results it can now give. Finally, we look at the expected future developments.

Background physics

An optical fibre can act as a light guide because there is total internal reflection at the boundary between a core with a high refractive index n_C and a sheath or 'cladding' with a low refractive index n_M . In principle it would also be possible to use a homogeneous optical fibre surrounded by air. However, in fibres for communications the surface where the total internal reflection occurs is moved inside the fibre because of the much lower attenuation and dispersion; we shall return to this later. The fibre is surrounded by a protective coating. Typical dimensions for a communications optical fibre are:

- core diameter 10 to 50 μm ,
- outer diameter of the cladding 125 μm and
- outer diameter of the protective coating 250 μm .

Important quantities that characterize the quality of an optical fibre are pulse broadening and attenuation. We shall start by considering the influence of these two factors.

Pulse broadening

A signal pulse is always broadened to some extent during transmission in the optical fibre, see fig. 3. The reason for this is dispersion, which can be of three different types: mode dispersion, material dispersion and waveguide dispersion.

We shall first consider mode dispersion, for the simplest type of optical fibre, with a stepped refractive-index profile ('step-index fibre'), see fig. 4a. In a fibre of this type the refractive index in the core is constant and only slightly higher than that of the cladding. The

stepped function of the radius. An optical fibre with a refractive-index profile of this form is called a 'graded-index fibre', see fig. 4b. The smallest value for the time-delay difference due to mode dispersion that can now be obtained with these fibres is 50 ps/km. The

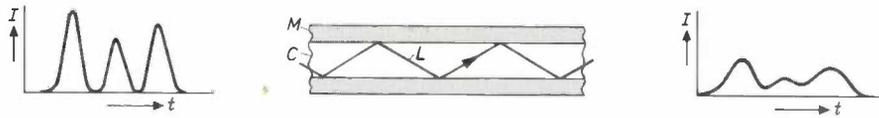


Fig. 3. Pulse broadening. Light pulses, shown here as the light intensity I as a function of time t , gradually become broader as the modulated light ray L propagates in an optical fibre with core C and cladding M .

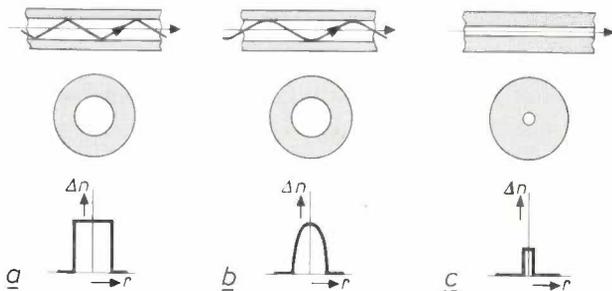


Fig. 4. a) Multimode fibre with stepped refractive-index profile. b) Graded-index multimode fibre with a continuous profile. c) Single-mode fibre. This fibre also has a stepped profile, but the core diameter is so small that a light ray can only propagate in a single mode. Top row: longitudinal sections. Centre row: cross-sections. Bottom row: refractive-index profiles, the difference Δn in refractive index between core and cladding as a function of the radius r . Dark grey: cladding.

refractive-index profile required for this must satisfy the equation

$$n(r) = n_{r=0} \left\{ 1 - \Delta n_{rel} \left(\frac{r}{r_C} \right)^\alpha \right\},$$

where r_C is the maximum radius of the core. Δn_{rel} refers to the largest relative difference in the refractive indices and α is the 'profile parameter'. The profile parameter has a value of about 2; its exact value depends on the wavelength λ of the light, Δn_{rel} and the elements used as dopants in making the glass. The product of the maximum signal frequency and the fibre length for these fibres is 10 to 20 GHz km in the ideal case.

The mode dispersion can be completely suppressed by reducing the core diameter until only one mode can exist, see fig. 4c. These optical fibres are called single-mode fibres and their index-profile is stepped. If fibres of this type are designed for a wavelength of 1300 nm, the core diameter is only 8 to 10 μm and the relative difference in refractive-index Δn_{rel} is 0.3 to 0.5%. In these fibres the maximum signal frequency is limited only by material dispersion and waveguide dispersion.

Material dispersion arises because the refractive index — and hence the velocity — is a function of the wavelength of the light in the material. This function depends on the composition of the glass. Since every light source has a finite spectral width, some pulse broadening due to material dispersion always occurs. Material dispersion is only of interest if mode dispersion is small.

relative difference Δn_{rel} in the refractive index, defined as:

$$\Delta n_{rel} = \frac{n_C - n_M}{n_C},$$

is about 1%. Inside the core the light from the light pulse can propagate in various modes, which are determined by the fact that constructive interference must occur. The difference in optical pathlength and hence in time delay for the different modes causes the pulse broadening. For a step-index fibre the difference in delay time is about 50 ns/km. This means that the product of the maximum signal frequency and the fibre length cannot be much more than 20 MHz km in such fibres. This figure is no higher than for coaxial cables.

The mode dispersion can be reduced by making the optical pathlength the same for the different modes. It is then necessary to make the refractive index vary as a continuous function $n(r) = f(r)$, rather than a

- [3] K. C. Kao and G. A. Hockham, Dielectric-fibre surface waveguides for optical frequencies, Proc. IEE 113, 1151-1158, 1966.
- [4] L. J. M. Bollen, J. J. Goedbloed and E. T. J. M. Smeets, The avalanche photodiode, Philips Tech. Rev. 36, 205-210, 1976.
- [5] F. P. Kapron, D. B. Keck and R. D. Maurer, Radiation losses in glass optical waveguides, Appl. Phys. Lett. 17, 423-425, 1970.

The last form of dispersion we shall consider here is waveguide dispersion, which again is significant in single-mode fibres. It arises because there is only one possible mode, in which the light wavefront propagates in the cladding as well as in the core. Waveguide dispersion depends on wavelength, composition of the material and the shape of the refractive-index profile. Fortunately, material dispersion and waveguide dispersion are of opposite sign, so that in single-mode fibres a certain amount of compensation between these two kinds of dispersion can be obtained by careful choice of dimensions and composition of core and cladding.

Fig. 5 shows how the theoretical product of maximum signal frequency and fibre length varies as a

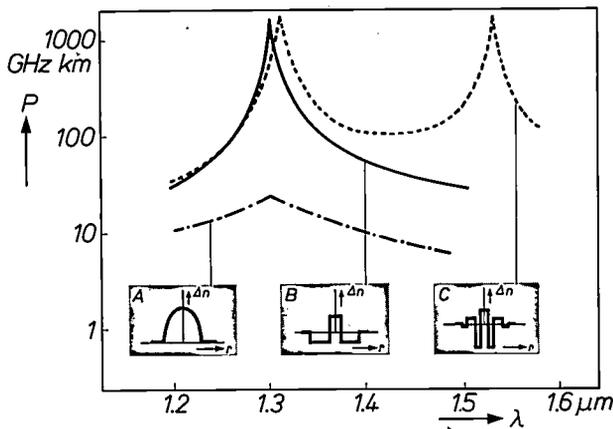


Fig. 5. The product P of maximum signal frequency and fibre length as a function of the wavelength λ of the light in the fibre. *A* curve for a graded-index multimode fibre. *B* curve for a single-mode step-index fibre. *C* curve for a single-mode fibre with double cladding, a DFSM fibre (Dispersion-Flattened Single-Mode). The curves are the result of measurements on fibres manufactured on a laboratory scale. Windows can clearly be distinguished at wavelengths of about 1300 and 1550 nm.

function of the wavelength for three well-known types of fibre. Curve *A* refers to a graded-index fibre, curve *B* to a single-mode fibre and *C* to a derivative, the particularly interesting DF single-mode fibre (DF stands for Dispersion-Flattened) or DFSM fibre. The bandwidth of the graded-index fibre (*A*) is limited owing to mode dispersion, material dispersion and waveguide dispersion. The bandwidth of the single-mode fibre (*B*) is larger, since mode dispersion has disappeared completely. The bandwidth of the DFSM fibre is the largest, because the specially shaped index profile gives compensation between waveguide dispersion and material dispersion. This compensation is par-

ticularly effective at wavelengths of about 1300 and 1550 nm.

Attenuation

The most important sources of signal attenuation in optical fibres are absorption and scattering. The decrease in the light intensity I as a function of the wavelength λ and the length x of the light path in the fibre is given by

$$I(\lambda, x) = I_0(\lambda) e^{-\alpha_{\text{TOT}} x}, \quad (1)$$

where $I_0(\lambda)$ is the intensity of the light at wavelength λ at the input to the fibre and $\alpha_{\text{TOT}}(\lambda)$ in m^{-1} is the loss factor due to both absorption and scattering. Fig. 6 shows the loss factor for a single-mode fibre in dB/km as a function of the wavelength, and the various components of the loss factor:

- scattering (chiefly Rayleigh scattering),
 - absorption at the short-wave end of the spectrum and
 - absorption at the long-wave end of the spectrum.
- The minimum value of α_{TOT} — about 0.20 dB/km — is at a wavelength of 1550 nm.

The limitation at the short-wave end of the 'transmission window' at 1550 nm is due to Rayleigh scattering and absorption, which occur in both pure and doped quartz glass. At the long-wave end the window is limited only by absorption, which is mainly due to the elements added to the quartz glass. The size and location of the window can be altered by varying the concentration of the 'dopants': GeO_2 , B_2O_3 , P_2O_5 and compounds that contain the element F. The minimum in the curve of fig. 6 generally becomes higher as a result of doping. The total attenuation can be lower than 0.2 dB/km if contamination by OH^- ions and metals can be avoided. Contamination by OH^- ions at a relative concentration by weight of only 10^{-6} increases the attenuation to 50 dB/km at a wavelength of 1380 nm. Even more serious is the effect of contamination by metal ions: contamination by Fe^{2+} ions at a thousandth of that concentration increases the attenuation by another 0.7 dB/km at a wavelength of 1100 nm.

The PCVD process

Initial requirements

A process for the mass production of high-quality fibres for optical communications must have the following features:

- exceptionally pure basic materials,
- no inhomogeneities such as schlieren or bubbles in the end-product,

- perfect control of the index profile in the radial direction,
- less than 1% axial variation of refractive index,
- high yield,
- high process output,
- high reproducibility and
- high degree of automation.

We shall now see how closely the Philips PCVD process can match this long list of requirements.

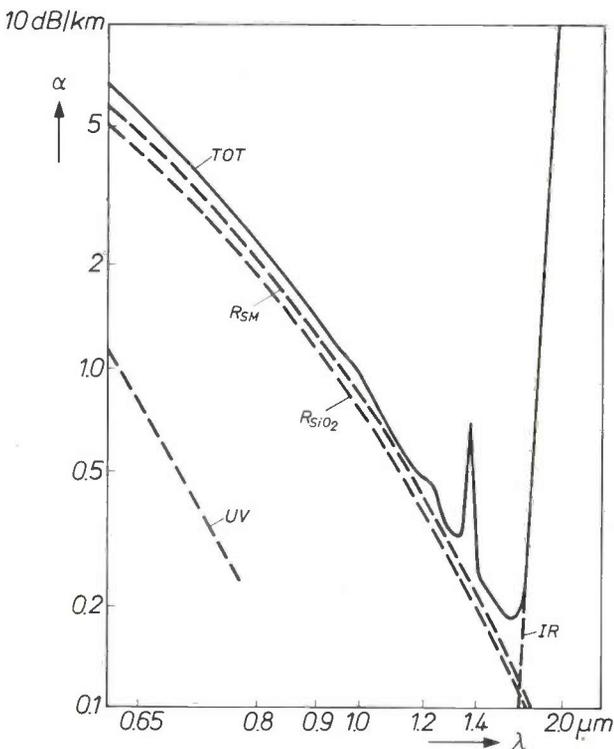


Fig. 6. The loss factor α due to attenuation as a function of the wavelength λ of the light. *UV* the attenuation due to absorption alone, at the short-wave end of the spectrum. *IR* the attenuation due to absorption alone, at the long-wave end. R_{SiO_2} the attenuation due to Rayleigh scattering in pure SiO_2 . R_{SM} the attenuation due to Rayleigh scattering in a single-mode fibre. All the above curves are theoretical. The curve labelled *TOT* is the result of measurements and refers to the total attenuation in a single-mode fibre. It is approximately equal to the sum of the curves *UV*, *IR* and R_{SM} .

The process

In spite of a great deal of effort, success in developing conventional glass-forming processes such as the double-crucible process^[6] into a mass-production technology for high-quality optical fibres has proved elusive. As mentioned earlier, the techniques that have been used successfully are all based on the CVD process: reactive deposition from the gas phase. The characteristic feature of this process is that glass con-

stituents such as SiO_2 and GeO_2 are formed from the gases $SiCl_4$, $GeCl_4$ and O_2 . These gases react with one another, and the reaction products condense on the walls of the enclosure in which the reaction takes place.

The PCVD process was proposed and developed by Philips Forschungslaboratorium Aachen in the early seventies^[7], and was later taken over by the Philips Glass Product Group^[8]. The process makes use of a plasma that stimulates the reactions between the various gases. Fig. 7 shows the arrangement. The most important components are the gas-supply system, the reaction zone inside a quartz-glass tube and the pump system.

In the gas-supply system the gases $SiCl_4$, $GeCl_4$, C_2F_6 and O_2 are metered with the aid of very accurate

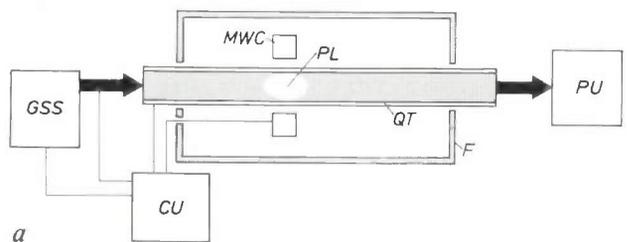


Fig. 7. a) Diagram of the PCVD process. *GSS* gas-supply system. *MWC* microwave cavity. *PL* plasma. *PU* pump system. *CU* control unit. *QT* quartz-glass tube. *F* furnace. b) The laboratory arrangement. The cavity can be seen on the left of the furnace insulation; it is white and encloses the quartz-glass tube.

[6] H. M. J. M. van Ass, P. Geittner, R. G. Gossink, D. Küppers and P. J. W. Severin, The manufacture of glass fibres for optical communication, Philips Tech. Rev. 36, 182-189, 1976.

[7] P. Geittner, D. Küppers and H. Lydtin, Low-loss optical fibres prepared by plasma-activated chemical vapor deposition (CVD), Appl. Phys. Lett. 28, 645-646, 1976.

[8] P. Matthijssse, The PCVD process for large scale production of telecommunication fibres, Proc. SPIE 584, 11-15, 1986.

mass-flow controllers, mixed and fed into the reaction zone through a common feed. The reaction zone is enclosed by the tube of pure SiO_2 , which functions as a substrate for the light-guiding material of the core and the cladding. Around the tube there is a furnace. The gas mixture is fed through the reaction zone to the pump system and passes through the plasma, where the gas molecules are fragmented into reactive particles.

The pressure in the reaction zone is 1 to 2 kPa (10 to 20 mbar). A ring-shaped microwave cavity resonator located around the SiO_2 tube excites the plasma. The frequency of the microwave energy is 2.45 GHz. The maximum power in the cavity is 6 kW. The reactive particles react in the plasma and the condensable reaction products are deposited on the wall of the relatively cold SiO_2 tube (about 1200 °C).

The special feature of the Philips PCVD process is that the cavity continually moves to and fro at high speed, taking the plasma with it inside the quartz-glass tube. A great many thin layers of glass form on the inside of the tube, distributed uniformly over the length. The thickness and the composition of any layer can be varied by varying the rate of travel of the cavity and the flow rate of the gases. Almost any index profile in 500 to 2000 extremely small steps can be obtained inside the SiO_2 tube in this way.

The useful part of the tube has a maximum length of 1 metre. The layers are constant in thickness and composition as a function of the length and the optical properties hardly vary at all. After the unwanted ends have been removed the tube is collapsed through the action of surface tension at a temperature of 2000 to 2200 °C to form a solid rod; see fig. 8. This rod, the 'preform', has a diameter of 15 to 30 mm and has the desired radial optical structure, but on a larger scale, see title photograph. The preform is next drawn to form an optical fibre several kilometres long; see fig. 8c and d. The fibre is finally given a plastic protective coating.

Significant advantages of the PCVD process are that

- the desired index profile can be approximated very accurately with such a large number of layers,
- the reaction and deposition rates are almost independent of the ambient temperature,
- yields are high,
- the structure on the inside of the SiO_2 tube is formed at relatively low temperatures,
- the energy in the cavity is transferred to the plasma with virtually no losses, regardless of the thermal characteristics of the tube and the furnace,
- the conditions hardly vary at all during the process, so that it is very suitable for the deposition of large

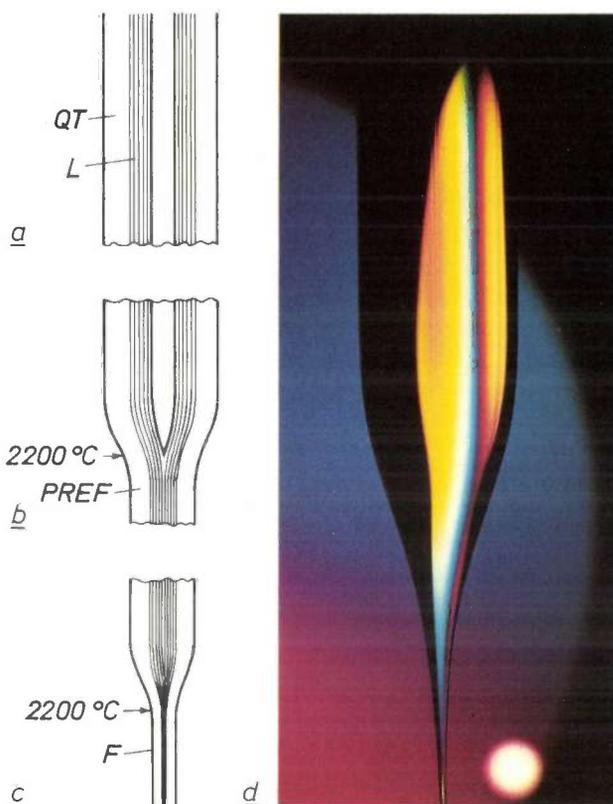


Fig. 8. The process steps that ultimately provide a glass fibre for optical communications. *a*) PCVD process. *QT* quartz-glass tube. *L* the layers deposited inside the tube by this process. *b*) Collapsing the tube to form a 'preform' *PREF*. *c*) Drawing the preform into a long thin glass fibre *F*. The regions at the highest temperature (2200 °C) are shown. *d*) Close-up of the 'neck' of the preform in the drawing process.

numbers of layers with a high combined thickness. The process is therefore suitable in principle for the manufacture of large preforms for very long fibres.

Current results

The good results that can now be achieved for attenuation, dispersion and process economics are largely due to:

- a good understanding of the factors that affect the deposition of SiO_2 and doping with GeO_2 and F,
- proper control of the index profile,
- the avoidance of contamination by OH^- and metal ions,
- effective control of the deposition rate, yield, reproducibility and dimensions of the preform.

Fig. 9 shows that the PCVD process will give relatively large variations in refractive index. These variations are larger than those that can be obtained with other processes. Only after we had gained a thorough knowledge of the application of the dopants in SiO_2 , and of the behaviour of these dopants as a function of temperature, was it possible to reproduce the desired

index profile sufficiently accurately in production [9]. This meant that by 1983 Philips were able to meet the specifications set by Deutsche Bundespost, the West German postal and telecommunications authority, for graded-index multimode fibres, and an order was received for large quantities of these fibres. The

why most of the research since the early eighties has been on single-mode fibres. It has also been found that the experience gained in the production of graded-index fibres is useful in the production of single-mode fibres.

Fig. 10 shows the index profiles that could be obtained in preforms before they were drawn. Although not all the profiles can be used in practice, the diagram clearly shows the rotational symmetries and radial accuracies that can be obtained [10]. Analyses of these profiles have indicated that the relative devia-

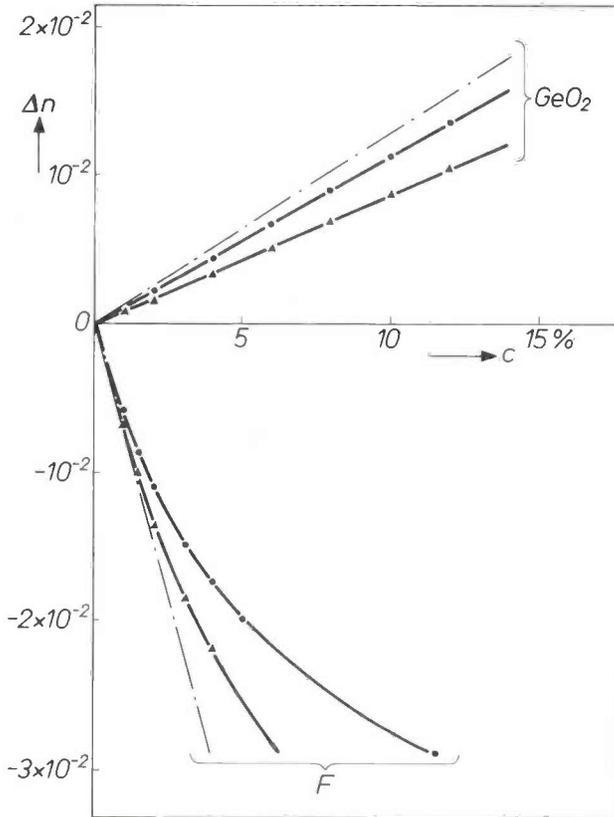


Fig. 9. The variation Δn in refractive index that can be obtained by introducing a particular molecular concentration c of dopant. The curves for positive values of Δn refer to GeO_2 molecules, the curves for negative values refer to the concentration of F atoms. The chain-dotted lines are the theoretically attainable values for complete incorporation of the dopants. The values that can be obtained in practice are indicated by the measured points and refer to the following process conditions: temperature 1200 °C, pressure 16 to 20 mbar, velocity of the microwave cavity 0.17 m/s, external diameter of the quartz-glass tube 26 mm, internal diameter 23 mm. For GeO_2 the deposition rate is 8 mg/s and the small filled triangles refer to a temperature of 1240 °C instead of 1200 °C. For fluorine the small filled circles correspond to a deposition rate of 8 mg/s and the small filled triangles to a rate of 25 mg/s.

length-bandwidth product for the fibres had to be at least 1.2 GHz km. At present the mean value for production graded-index fibres is 2.5 to 3 GHz km. For 6% of the fibres the value is as high as 4 GHz km.

Although these results are very satisfactory, it was soon clear that single-mode fibres for the wavelength range from 1300 to 1600 nm do have advantages over graded-index fibres in bit rate and attenuation. This is



Fig. 10. Various measured 3-dimensional index profiles in the preforms used in the drawing process. The radial scales are not the same in all the diagrams. The diagram at the lower right corresponds to the DFSM fibre (profile C in fig. 5). When fibres are drawn from the preforms the shape of their index profiles is virtually unchanged.

[9] P. Geitner, H. Hübner, H. Lydtin and H. Wilson, Influence of central profile distortions on the bandwidth of GRIN multimode fibres, Proc. Conf. Opt. Fiber Commun. WG2, 90-91, New Orleans 1983.

[10] H. Lydtin, PCVD: a technique suitable for large-scale fabrication of optical fibers, J. Lightwave Technol. LT-4, 1034-1038, 1986.

tions from the ideal profile are smaller than 1%. The profile at the lower right refers to a preform for the DFSM fibre mentioned earlier. In a wide wavelength range from 1320 to 1600 nm this type of fibre has an extremely low time-delay difference due to dispersion, in relation to the spectral width of the light source: about $2 \text{ ps nm}^{-1} \text{ km}^{-1}$. DFSM fibres have in the meantime been included in the Philips product range.

Fig. 11 gives an impression of the spread in the attenuation at a wavelength of 1310 nm in fibres made

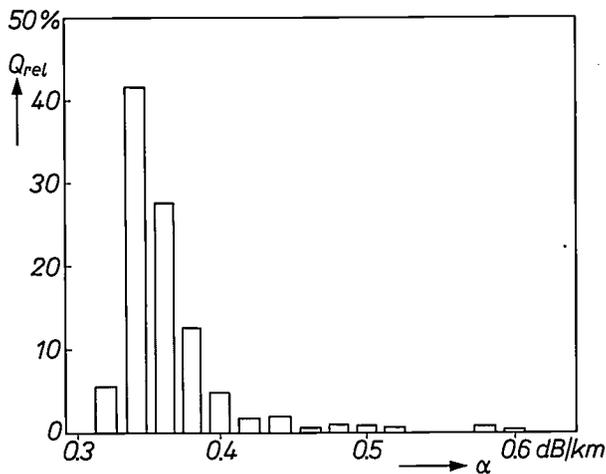


Fig. 11. Bar diagram showing the relative quantity Q_{rel} of drawn fibre as a function of the measured loss factor α due to attenuation at a wavelength of 1310 nm. The diagram gives an impression of the production spread in the current manufacturing process for single-mode fibres.

by the present production process. The mean attenuation of 0.35 dB/km is only 0.05 dB/km more than the minimum attainable value. The most important condition for obtaining low attenuation is the reduction of the contamination by OH^- ions. Until recently this was a serious problem: investigations had shown that in the PCVD process one in 80 of the hydrogen atoms in the reactive gas mixture was taken up by the layer structure as an OH^- ion. In the early eighties this high 'take-up efficiency' led to increases in the attenuation of more than 100 dB/km at a wavelength of 1380 nm. This made it impossible to use glass optical fibres in the 1300-nm transmission window. The problem of OH^- -ion contamination was solved by adding fluorine-containing gases such as C_2F_6 to the reactive gas mixture. Fig. 12 shows the decrease in the attenuation at 1380 nm as a function of the concentration of C_2F_6 in the gas mixture^[11]. The extra attenuation due to OH^- -ions at 1380 nm has recently been reduced to a minimum value of 0.1 dB/km.

For a total evaluation of the PCVD process it is not only necessary to consider the physical properties of the manufactured fibres — it is also desirable to keep the economic aspects in mind. After it had become clear that the process was suitable for manufacturing fibres with special index-profiles (see fig. 10), it was possible to show that the improved understanding of the process had led to an improvement in the reproducibility. This opened the way to the automation of the manufacturing process. It was also possible to show that deposition rates of the material in the quartz-glass tube of up to 3.3 grams per minute are possible^[12].

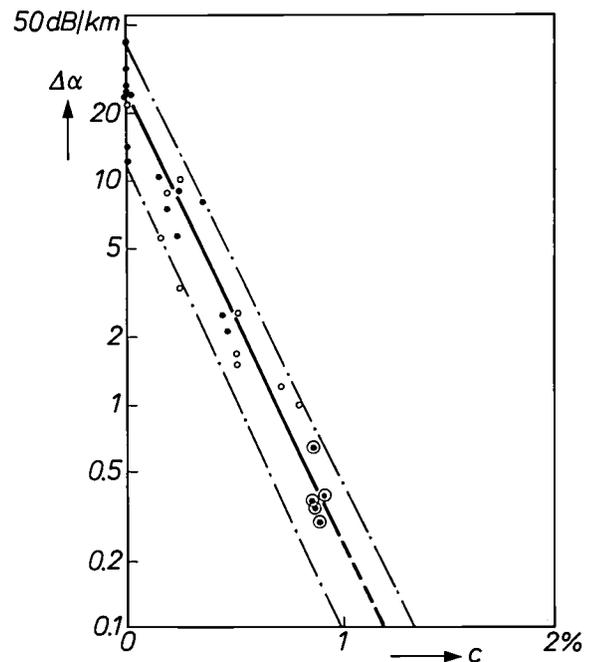


Fig. 12. The reduction $\Delta\alpha$ in the loss factor due to attenuation at a wavelength of 1380 nm as a function of the molecular concentration c of C_2F_6 in the reactive gas mixture in the PCVD process. The measured values shown as open circles correspond to single-mode fibres and the filled circles to graded-index multimode fibres. The reduction in the attenuation is due to the reduction in the numbers of OH^- ions in the preform^[11].

Outlook

Our future efforts will be concentrated on the continuous improvement of the PCVD process, with the main aim of manufacturing preforms equivalent to a longer fibre^[13]. The possibility of starting with a tube of high-purity quartz glass for the light-guiding part of the fibre cladding is also being investigated. The application of two tubes, one for the light-guiding

part and one for the non-light-guiding part of the cladding, saves time in the PCVD process and in collapsing the tubes. The larger preform that is the result of this improved process will give a fibre length of about 300 km per preform. So there is some evidence that the preparation of the core and the light-guiding cladding material will be separated.

There is a clear interest in single-mode fibres, since their characteristics are much better than those of the other types of optical fibres. Within the 'family' of single-mode fibres there is a noticeable trend towards

glass fibres with a complicated structure, in which various kinds of dispersion are compensated. An example of this is the DFSM fibre mentioned earlier, which can be manufactured relatively easily by the PCVD process.

Summary. Optical fibres for telecommunications are either multi-mode fibres (step- or graded-index) or single-mode fibres. Single-mode fibres have either a stepped profile with a small core diameter or a complex layered refractive-index profile. The last type of fibres are called DFSM (Dispersion-Flattened Single-Mode) fibres. In fibres of this type pulse broadening due to mode dispersion does not occur and material dispersion and waveguide dispersion almost completely compensate one another. The Philips PCVD (Plasma-activated Chemical Vapour Deposition) process can be used for the manufacture of fibres with complicated index profiles. Closely controlled amounts of dopants such as GeO_2 can be introduced into the basic SiO_2 material at very low concentrations. The concentration of OH^- ions, which can introduce an intolerably high attenuation, has been greatly reduced by adding C_2F_6 to the reactive gas mixture. This reduced the attenuation due to OH^- ions to a minimum value of 0.1 dB/km.

[11] P. Bachmann, P. Geittner, D. Leers, M. Lennartz and H. Wilson, Low OH excess loss PCVD fibres prepared by fluorine doping, *Electron. Lett.* **20**, 35-37, 1984.

[12] P. Geittner, H.-J. Hagemann, J. Warnier and H. Wilson, PCVD at high deposition rates, *J. Lightwave Technol.* **LT-4**, 818-822, 1986.

[13] P. Geittner, H.-J. Hagemann, H. Lydtin and J. Warnier, Hybrid technology for large SM fiber preforms, *J. Lightwave Technol.* **LT-6**, 1451-1454, 1988.

Phase-change optical recording

D. J. Gravesteijn, C. J. van der Poel, P. M. L. O. Scholte and C. M. J. van Uijen

Optical recording has been an important field of research at Philips Research Laboratories since the early seventies. This research has led to well-known applications such as 'Laser-Vision' and the 'Compact Disc'. Systems have also been developed in which the user can record information on the disc himself; one example of this is Digital Optical Recording (DOR) for data storage. There is also great interest in recording systems that allow both recording and erasure, e.g. by making use of magneto-optic effects or crystalline/amorphous phase change. While erasable magneto-optical recording has now reached a highly advanced stage of development, phase-change optical recording is still at the research stage. There have been so many recent developments, however, that an article in this journal on this method of optical recording now seems appropriate.

Introduction

Various systems for optical recording have been developed in the last fifteen years, and the distinctions between them relate to the applications and the recording media used. The best-known systems, 'Laser-Vision' [1] and 'Compact Disc Digital Audio' [2], are only intended for playback of the information stored on a disc. The information signal is recorded in the form of pits on the disc during manufacture. A laser beam is used in the playback process, which is based on the modulation of the reflection by the pits.

Optical recording systems have also been developed in which the user can record information on a disc himself. The player used here has a focused laser beam whose intensity is modulated to correspond to the information to be recorded. The disc has a film of material sensitive to this laser light. The laser radiation produces local effects in this film, and these effects give a satisfactory optical contrast on read-out with a laser beam (a much weaker one). Various kinds of effects can be used:

- the formation of holes by melting away the material, as in Digital Optical Recording (DOR) and in recording with organic dyes [3];
- local changes in the direction of magnetization in a

magnetic film subjected to an external magnetic field (magneto-optical recording), which can also be used to erase the recorded information [4];

- changes from the crystalline phase to the amorphous phase and vice versa, again with the possibility of erasure.

In optical recording in a polycrystalline film a short pulse with a focused laser beam makes the temperature locally higher than the melting point of the material, so that the area heated by the beam melts. After the pulse, the molten area cools down to room temperature so quickly that the disordered liquid state is 'frozen in' and no significant crystallization can occur. In this way an amorphous area is produced in a crystalline environment. Since the optical properties (refractive index, absorptance) of the amorphous phase are different from those of the crystalline phase, the recorded information can be read out as an optical contrast. The recorded information can be erased by heating the material again with a laser beam, but now to a temperature below the melting point. The amorphous area then recrystallizes because of the increased mobility of the atoms, and it can rapidly return to the thermodynamically more stable crystalline state (typically in less than 1 μ s).

Certain materials with a stable amorphous phase at low temperatures crystallize so rapidly on heating that

Dr D. J. Gravesteijn, Dr C. J. van der Poel, Dr P. M. L. O. Scholte and Dr C. M. J. van Uijen are with Philips Research Laboratories, Eindhoven.

it is also possible to use a completely amorphous film as the 'recording layer' for optical recording. The pulsed laser beam then forms crystalline areas in an amorphous environment. In this method of recording, however, it is not feasible to erase the information under practical conditions: after a crystalline area has been melted by brief laser heating there is too much recrystallization, in spite of the rapid cooling to room temperature.

The research described in this article was mainly concerned with the search for materials for erasable phase-change optical recording. The optical properties of the materials give a good indication of their potential usefulness, of course. The materials should be fusible at moderate laser powers (say 10 mW), and the amorphous areas produced should give sufficient contrast on read-out. The crystallization behaviour at various temperatures is also very important. If the recorded information is to be stable, the amorphous areas should have a long life at ambient temperatures as high as 70 °C, while for satisfactory erasure they should recrystallize rapidly (in less than 1 μ s) at several hundred °C. These last two requirements are a particular problem in the search for suitable materials.

A tellurium recording layer will meet most of the requirements for erasable optical recording, but the amorphous phase is not stable at room temperature. If an alloy such as $\text{Te}_{0.80}\text{Se}_{0.10}\text{Sb}_{0.10}$ is used instead of tellurium^[5], the stability of the amorphous phase at room temperature can be considerably increased, but because of the slower crystallization at higher temperatures, erasure is too slow. Materials such as the III-V semiconductors InSb and GaSb have a stable amorphous phase at room temperature, but crystallize so rapidly at high temperatures that they cannot be used for writing information into a crystalline film and are therefore unsuitable for erasable optical recording. Because of their rapid crystallization and the excellent stability of the amorphous phase at low temperatures, however, they can be used for writing information into an amorphous film for non-erasable recording^[6]. By adding a little Te to InSb and changing the ratio of In to Sb, an alloy can be obtained with the desired properties for erasable optical recording^[7].

In this article we shall first give a short description of the recording system, in which the most important aspects of writing, read-out and erasure are discussed. We next consider possible materials for the recording layer, with particular attention to the crystallization behaviour and the resulting effects on the choice of materials. Then we look more closely at experiments on non-erasable and erasable optical recording, and finally we shall give some general indications of the prospects for practical application.

The recording system

As in other forms of optical recording the recording system consists of three main components: an optical disc, a motor that rotates the disc at the correct speed, and an arm with a 'light pen'^[8]. The disc consists of a light-transmitting pregrooved substrate, a light-sensitive recording layer and a protective coating. The operation of the light pen is shown schematically in *fig. 1*; the example shown is from a Compact Disc

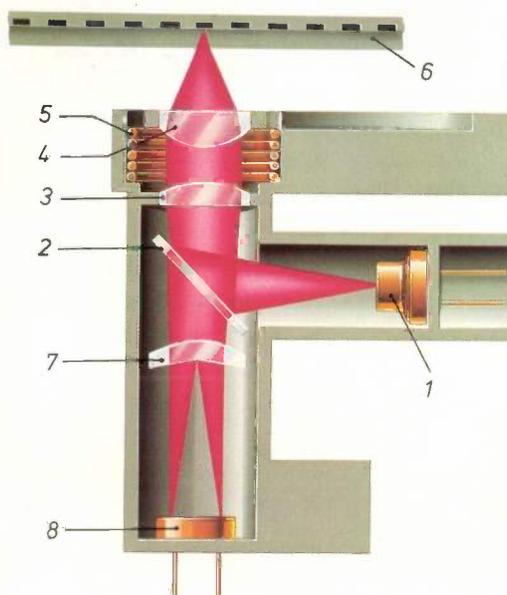


Fig. 1. Diagram of the operation of the light pen in a Compact Disc player. 1 semiconductor laser. 2 half-silvered mirror. 3 collimator lens that intercepts the laser light. 4 objective lens for focusing the light beam. 5 actuator for positioning the objective. 6 optical disc. 7 beam-splitter. 8 detector for the light reflected from the disc.

- [1] H. C. Haverkorn van Rijsewijk, P. E. J. Legierse and G. E. Thomas, Manufacture of LaserVision video discs by a photopolymerization process, *Philips Tech. Rev.* **40**, 287-297, 1982.
- [2] Special issue on the Compact Disc, *Philips Tech. Rev.* **40**, 149-180, 1982.
- [3] L. Vriens and B. A. J. Jacobs, Digital optical recording with tellurium alloys, *Philips Tech. Rev.* **41**, 313-324, 1983/84; D. J. Gravesteijn and J. van der Veen, Organic-dye films for optical recording, *Philips Tech. Rev.* **41**, 325-333, 1983/84.
- [4] M. Hartmann, B. A. J. Jacobs and J. J. M. Braat, Erasable magneto-optical recording, *Philips Tech. Rev.* **42**, 37-47, 1985.
- [5] C. J. van der Poel, D. J. Gravesteijn, W. G. V. M. Rippens, H. T. L. P. Stockx and C. M. J. van Uijen, Phase-change optical recording in TeSeSb alloys, *J. Appl. Phys.* **59**, 1819-1821, 1986.
- [6] D. J. Gravesteijn, H. M. van Tongeren, M. Sens, T. Bertens and C. J. van der Poel, Phase-change optical data storage in GaSb, *Appl. Opt.* **26**, 4772-4776, 1987.
- [7] D. J. Gravesteijn, Materials developments for write-once and erasable phase-change optical recording, *Appl. Opt.* **27**, 736-738, 1988.
- [8] See for example G. E. Thomas, Future trends in optical recording, *Philips Tech. Rev.* **44**, 51-57, 1988.

player. A modulated AlGaAs laser gives a divergent light beam at a wavelength of about 800 nm. The beam is incident on a half-silvered mirror and the reflected light is intercepted by a collimator lens. An objective lens with a numerical aperture of 0.4 to 0.5 focuses the beam on the recording layer on the disc. This results in the read-out, writing or erasure of information, depending on the laser power and the duration of the pulse.

The information density that can be achieved is determined by the 'effective diameter' of the light spot on the disc. Because of diffraction this diameter cannot be less than about $0.5\lambda/NA$, where λ is the laser wavelength and NA the numerical aperture of the objective lens. If $\lambda = 800$ nm and $NA = 0.4$ the spot has an effective diameter of about 1 μm .

The light reflected by the disc enters the detector branch of the light pen and is converted into electrical signals. These are used for generating tracking signals, which ensure that the laser beam remains focused on the disc and follows the correct track. They are also used for generating synchronization signals and providing address information. In the read-out process the difference in reflectance between crystalline and amorphous areas is also detected. This provides the high-frequency signal containing the stored information in the recording layer.

The laser beam is focused through the substrate on to the recording layer. The substrate is thick enough (about 1 mm) to keep dust and damage to the surface well outside the depth of focus of the objective, so that they are no longer a problem.

We shall now look more closely at the more important aspects of the writing, read-out and erasure of information.

Writing

Information can be written into a *polycrystalline* film (fig. 2a) with a pulsed laser beam, with a pulse duration of say 100 ns and a laser power of say 10 mW at the disc. The laser energy locally incident on the disc is partly absorbed by the recording layer. This produces an increase in temperature to above the melting point of the material, so that the material melts locally. After the pulse the molten material cools rapidly, at cooling rates reaching 10^8 to 10^9 K/s. Because of the rapid cooling the material does not have sufficient time to crystallize and the disordered liquid structure is 'frozen in'. In this way an amorphous area is produced in a crystalline environment. The diameter of this area is about 1 μm , corresponding to the effective diameter of the laser spot on the disc.

In writing in an *amorphous* recording layer (fig. 2b) the laser power and the pulse duration are given val-

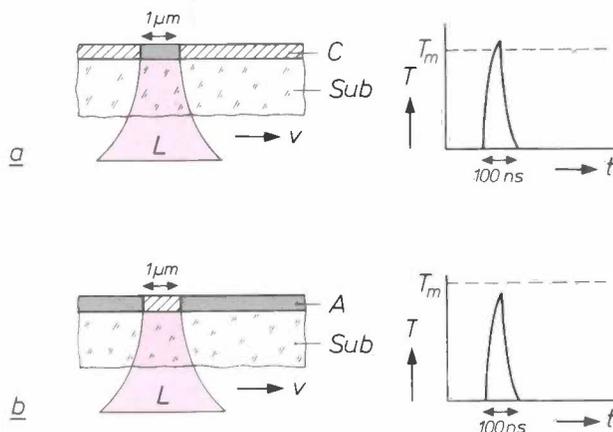


Fig. 2. a) Principle of writing information into a crystalline film, the recording layer C , on an optical disc rotating at a linear velocity v . During a pulse of say 100 ns a laser beam L is focused through the substrate Sub on to the crystalline film. The laser beam heats an area of diameter about 1 μm , whose temperature T rises to above the melting point T_m of the material. After the pulse this area cools so rapidly that it becomes amorphous. b) Principle of writing in an amorphous recording layer A . The material is heated to a temperature below the melting point. The area heated by the beam becomes completely crystalline.

ues such that the temperature of the heated material remains below the melting point. This gives local crystallization of an area with a diameter of about 1 μm .

Read-out

The information can be read out (fig. 3) with a continuous-wave (c.w.) laser beam that scans the track carrying the recorded information. Here the power in the beam should be low (1 mW, say), so that it does not introduce phase changes in the recording layer, since this would distort the recorded information. The light reflected through the disc is detected in the scanning process. Because of the difference in reflectance between the amorphous and crystalline phases the intensity of the reflected light is modulated: more light or less light is reflected depending on the position of the beam on the disc. This gives an intensity that varies with time, corresponding to the recorded information.

Erasure

The erasure of information written into an originally polycrystalline recording layer is illustrated schematically in fig. 4. An area that has become amorphous in the writing process can be made crystalline again by heating it with the laser beam to a temperature below the melting point of the material. The time required for the crystallization must be shorter than the 'pass time' of the laser beam, i.e. the time during which the area is heated. Good results have been obtained with an appropriate choice of material and a pass time of about 1 μs . If materials with a longer

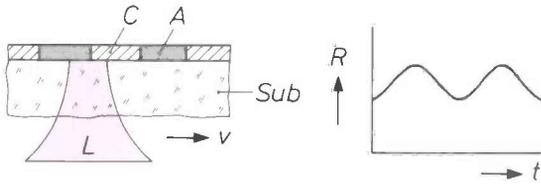


Fig. 3. Principle of the read-out of written information. A low-power continuous-wave (c.w.) laser beam L scans the recording layer and the reflected light is intercepted by the detector (fig. 1). The difference in reflectance between the crystalline areas C and the amorphous areas A produces an intensity variation of the reflected light R with time t .

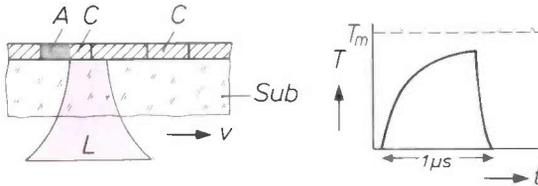


Fig. 4. Principle of the erasure of written information. A focused laser beam with a pass time of say $1 \mu\text{s}$ gives local heating to a temperature T below the melting point T_m . Areas that had become amorphous in the writing process are made crystalline again by this heating.

crystallization time are used it may be desirable to increase the pass time. This can be done by using a lengthened spot, but a separate erase laser with a different optical system is then required.

Materials for the recording layer

The material of a recording layer for phase-change optical recording has to meet so many different requirements that the choice of material is limited. The material must be one that can be applied as a thin optically homogeneous film, e.g. by evaporation or magnetron sputtering. It must be possible to introduce phase changes in the material by heating it with the laser beam. This means that it must absorb sufficiently at a wavelength of about 800 nm, the emission wavelength of the laser. For read-out of the written-in 'marks' the material must also reflect sufficiently at the same wavelength, and there must be sufficient optical contrast between the amorphous and crystalline phases. To limit the noise level the crystalline phase must only contain small crystallites. It must be possible to repeat the read-out many times (say 10^8 times) without any change in the material. And it must also be possible for both phases to preserve their structure and composition for a long time, say ten years in an ambient temperature of about 70°C .

For erasable optical recording there are a few extra requirements: the material must be capable of being switched many times ($> 10^6$) between the amorphous

and crystalline states, and the written effects should be rapidly erasable, preferably in less than $1 \mu\text{s}$. The suitability of a material for erasable recording is largely determined by its crystallization behaviour at various temperatures.

Crystallization behaviour

The crystallization of an amorphous material is usually described as a two-stage process: nucleation, in which small nuclei of crystalline material are formed, and the growth of these nuclei to form crystallites^[9]. In both stages the reduction in the free energy is the 'driving force'. The nucleation can occur anywhere in the interior of the material (homogeneous nucleation) or at impurities and boundary surfaces (heterogeneous nucleation). We shall only consider homogeneous nucleation here.

The formation of a crystalline nucleus of radius r has two opposing effects on the energy of the material: a reduction because the crystalline phase has a lower Gibbs free energy than the amorphous phase, and an increase because the surface energy increases with the increase in surface area. The change in energy ΔE is given by:

$$\Delta E = -\frac{4}{3}\pi r^3 \Delta G + 4\pi r^2 \sigma, \quad (1)$$

where ΔG is the difference in Gibbs free energy per unit volume and σ is the surface energy per unit area.

The variation of ΔE with r is shown schematically in fig. 5. Because of the opposing effects ΔE has a maximum, at a radius of $r_c = 2\sigma/\Delta G$. A nucleus with $r < r_c$ not only has a higher energy than the amorphous phase but will not grow very easily, since an increase in r would make the energy difference from the amorphous phase even larger. A nucleus with $r > r_c$, on the other hand, can gain energy by growing. If they are to survive, the nuclei must therefore first reach the critical radius r_c and for this the energy barrier ΔE_{max} must be overcome. Substituting r_c in eq. (1) gives the following expression for ΔE_{max} :

$$\Delta E_{\text{max}} = \frac{16\pi\sigma^3}{3(\Delta G)^2}. \quad (2)$$

The number of suitable nuclei (i.e. with $r \geq r_c$) formed per unit time (the nucleation frequency), depends on the temperature. If the temperature is increased the energy barrier ΔE_{max} becomes higher because ΔG becomes smaller:

$$\Delta G = \Delta H_m - T\Delta S_m \approx \Delta H_m(1 - T/T_m), \quad (3)$$

^[9] See for example J. W. Christian, Phase transformations, in: Physical metallurgy, R. W. Cahn (ed.), North-Holland, Amsterdam 1970, pp. 479-486.

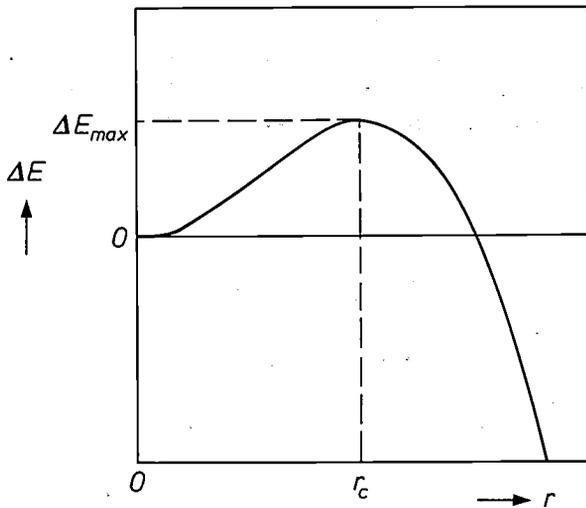


Fig. 5. Diagram showing the energy change ΔE on the formation of a crystalline nucleus from the amorphous phase, plotted against the radius r of the crystalline nucleus; see eq. (1). A decrease in the energy on further growth only occurs when the nucleus has reached a critical radius r_c , but the energy barrier ΔE_{max} must be overcome first.

where ΔH_m and ΔS_m are the enthalpy and entropy of fusion (per unit volume) and T_m is the melting point. On the other hand the mobility of the atoms is higher, which means that the viscosity of the material is lower. These opposing effects lead to the following expression for the nucleation frequency I :

$$I = f(T) \exp \{ -\Delta E_{max}(T)/kT \}, \quad (4)$$

where $f(T)$ is a monotonically increasing function of the temperature, which also contains the temperature-dependence of the viscosity. A temperature increase from room temperature at first gives a sharp rise in I , mainly as a result of the increasing mobility of the atoms. Above a certain temperature, however, ΔE_{max} becomes so large that I falls with temperature.

In the growth of the nuclei to form crystallites there are also the opposing effects of an increasing mobility and a decreasing driving force at higher temperatures. The following expression has been derived for the growth rate u ^[10]:

$$u = g(T) \left[1 - \exp \left\{ \frac{\Delta H_m}{kT} (1 - T/T_m) \right\} \right], \quad (5)$$

where the function $g(T)$ again contains the viscosity.

For a crystallization process consisting of homogeneous nucleation and three-dimensional growth the fraction $x(t)$ that has crystallized at time t at a constant temperature can be written as^[9]:

$$x(t) = 1 - \exp(-\pi I u^3 t^4/3), \quad (6)$$

where I and u are given by eqs (4) and (5).

We have used this expression to calculate the time required to crystallize a particular fraction (say 2/3) of an area of diameter 1 μm in an amorphous tellurium film. The result is shown in fig. 6 (curve *a*). The calculated crystallization time first decreases rapidly with temperature and then increases rapidly again as a result of the reduction of the nucleation frequency and the growth rate. In the entire temperature range,

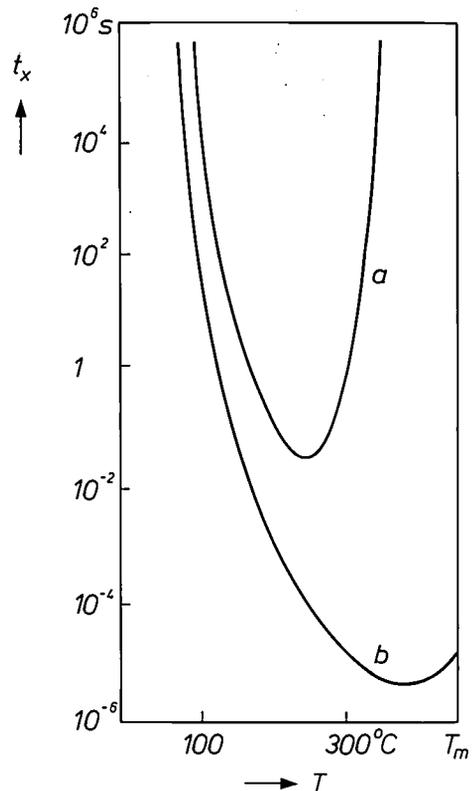


Fig. 6. Calculated temperature dependence of the time necessary for tellurium to crystallize throughout an area of diameter 1 μm . Curve *a* was calculated for homogeneous nucleation followed by three-dimensional growth, with the value of $x(t)$ in eq. (6) taken as 2/3. As the temperature T increases t_x first falls sharply, then rapidly rises again. In calculating curve *b* it was assumed that the only rate-determining factor was the growth of the crystallites from crystalline boundaries. The rapid fall in t_x is reversed at a temperature just below the melting point T_m .

however, the calculated times are many orders higher than the minimum crystallization times measured experimentally (0.2 μs).

Better agreement is obtained if it is assumed that the only rate-determining factor is the growth inwards from the crystalline boundaries for a distance of 0.5 μm . In this case the calculated crystallization time continues to decrease rapidly to a temperature just below the melting point (curve *b* of fig. 6) and the minimum crystallization time is much closer to the experimental value. The values for the crystallization

time in the neighbourhood of room temperature, which are of interest for the stability of written information, are unreliable, since the viscosity data required are not known and can only be estimated by a rough extrapolation from higher temperatures. The crystallization time for tellurium at room temperature is less than 1 s, much too short for practical applications.

Choice of materials

The discussion above gives some indication of the materials that might be used for the optical recording layer. For erasable recording the minimum crystallization time t_x^{\min} must be low enough. In general a small structural difference between the amorphous and crystalline phases will have a beneficial effect on t_x^{\min} , but will be a disadvantage for the stability of the amorphous phase at room temperature. The stability can be increased by alloying with other elements, but this makes t_x^{\min} higher. This has been found, for example, with the addition of Se and Sb to tellurium^[6]. In TeSeSb alloys with an amorphous phase of sufficient stability (ten years or more at room temperature), t_x^{\min} has a value of more than 10 μ s.

A shorter crystallization time can be obtained by taking a material with a higher value of ΔH_m , so that ΔG becomes larger; see eq. (3). But this makes the stability at room temperature worse. This can be corrected by using a material with a higher melting point. The supercooling of amorphous areas at room temperature is then greater, so that the atoms are less mobile.

Materials with the desired crystallization properties are found in the class of compounds consisting of elements from Groups III and V of the periodic table. If the requirements of sufficient absorptance, reflectance and optical contrast at the laser wavelength are also considered, InSb and GaSb are found to be suitable candidates. In Table I the important properties of

Table I. Properties of interest for the crystallization behaviour of materials that can be used for phase-change optical recording. The table shows the melting point T_m , the enthalpy of fusion ΔH_m , the measured minimum crystallization time t_x^{\min} of amorphous material and the crystallization temperature T_x measured at a heating rate of 10 K/min.

Material	T_m (K)	ΔH_m (kJ/mol)	t_x^{\min} (ns)	T_x (K)
Te	723	18	200	270
InSb	800	48	15	460
GaSb	985	62	<7	620

these materials for crystallization behaviour are compared with those of tellurium. The materials have much higher ΔH_m -values and higher melting points. The measured minimum crystallization times are less than a tenth of that of tellurium, yet their amorphous phases at room temperature are much more stable. This is found from differential scanning calorimetry (or DSC), in which amorphous materials are heated at a constant heating rate of 10 K/min and the heat dissipated is measured as a function of the temperature. The temperature at which a sharp peak arises because of the occurrence of crystallization — the crystallization temperature T_x — is much higher for sputtered InSb and GaSb than for evaporated tellurium.

In the case of evaporated InSb the temperature dependence of the crystallization time was monitored in a wide range^[11]. The result, shown in fig. 7, shows that with materials of this kind the theoretically predicted very large variation of the crystallization time with temperature is indeed observed experimentally.

The materials InSb and GaSb are unsuitable for erasable optical recording, however, since their crys-

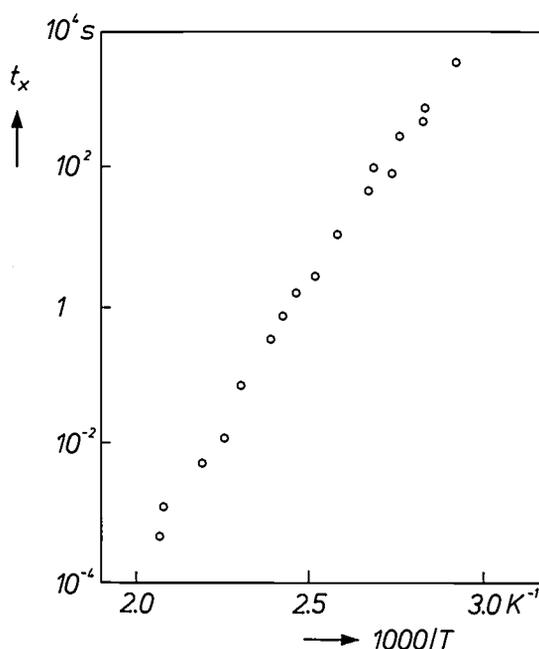


Fig. 7. Measured crystallization time t_x of evaporated InSb, plotted on a logarithmic scale against the reciprocal of the absolute temperature T . In the temperature range investigated the crystallization time varies by nearly eight orders of magnitude.

[10] D. R. Uhlmann, A kinetic treatment of glass formation, *J. Non-Cryst. Solids* 7, 337-348, 1972.

[11] C. J. van der Poel, Rapid crystallization of thin solid films, *J. Mater. Res.* 3, 126-132, 1988.

tallization rates at high temperatures are so high that once they have become crystalline they cannot be made amorphous again by the laser. The cooling after laser melting is too slow to prevent crystallization. The speed of the crystallization is demonstrated in *fig. 8*, a photograph made with a transmission electron microscope (TEM) after pulsed heating of an amorphous area between two crystalline tracks. During the pulse and the subsequent cooling a crystallite with a length of about $1\ \mu\text{m}$ has grown from the crystalline boundary in about 100 ns. This means that the growth rate was about 10 m/s.

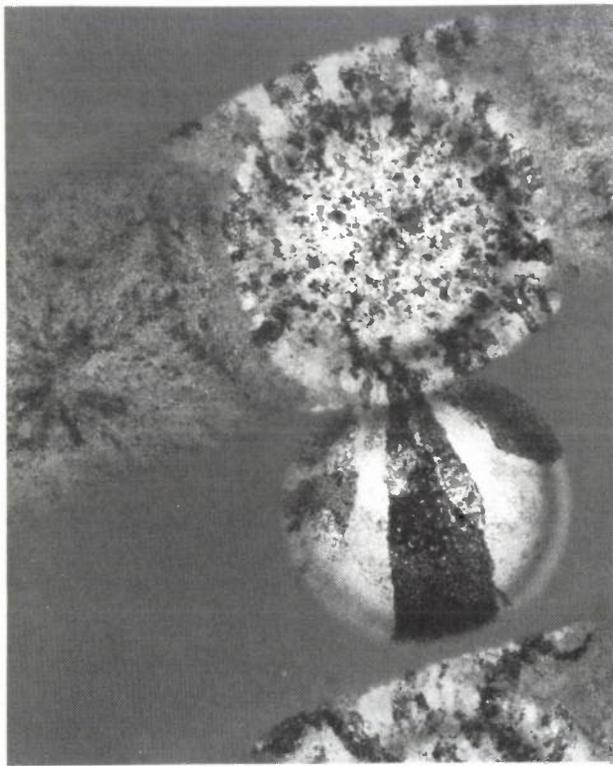


Fig. 8. TEM photograph of an amorphous area between two crystalline tracks after heating with a 60-ns write pulse. A crystallite about $1\ \mu\text{m}$ long has grown from the crystalline boundary.

For non-erasable optical recording in originally amorphous material GaSb is a particularly suitable material^[6]. The amorphous phase is sufficiently stable at room temperature and it crystallizes sufficiently rapidly on heating to just below the melting point. Adding 5 to 10 at. % Se or Te to InSb and changing the In/Sb ratio will slow down the crystallization at high temperatures sufficiently to permit erasable recording^[7]. Some results of experiments with non-erasable and erasable optical recording will now be briefly described.

Non-erasable optical recording

Amorphous GaSb films for non-erasable optical recording have been deposited on substrates such as glass and polycarbonate by magnetron sputtering from a target of pressed GaSb powder^[6]. The optimum film thickness can be derived by comparing the reflectance of amorphous films of different thicknesses with the reflectance of corresponding crystalline films. *Fig. 9* shows the reflectance of the amorphous films (R_a) and of the crystalline films (R_c) plotted against the film thickness. The variation of the optical contrast, defined as $|R_a - R_c|/(R_a + R_c)$, is also shown. The contrast is highest at a thickness of about 100 nm, the optimum film thickness for optical recording.

The result of a writing process in an amorphous GaSb film 100 nm thick can be seen in *fig. 10*. The areas where a phase change has occurred can easily be distinguished. The patterns obtained by local electron diffraction of the amorphous material and a written area are also shown. The diffraction pattern of the original material contains broad bands, indicating the presence of amorphous material. The pattern due to the written area, on the other hand, contains sharp diffraction rings with dots, characteristic of a polycrystalline material; the position of the diffraction rings corresponds to the lattice constants of GaSb.

Recording experiments have been performed with discs consisting of a pregrooved polycarbonate substrate, a 100-nm layer of amorphous GaSb and a protective coating. The objective lens used in the recording equipment had a numerical aperture of 0.45 and the laser operated at 780 nm. Information was written

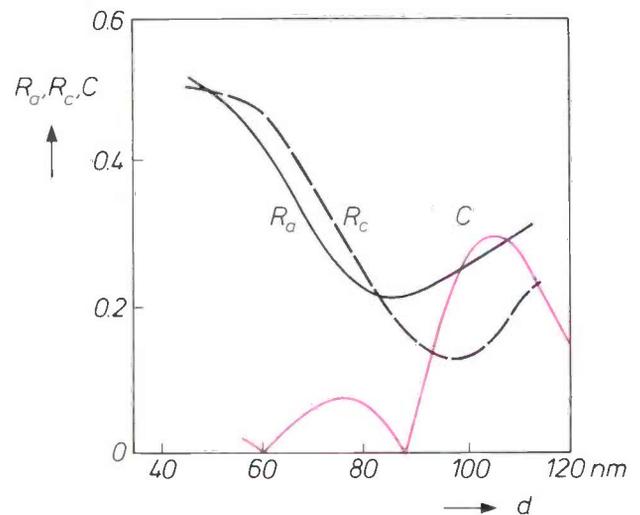


Fig. 9. Reflectance of GaSb films as a function of the thickness d for the amorphous phase (R_a) and the crystalline phase (R_c), and the optical contrast C , defined as $|R_a - R_c|/(R_a + R_c)$, derived from these values. The contrast is highest at a film thickness of about 100 nm.

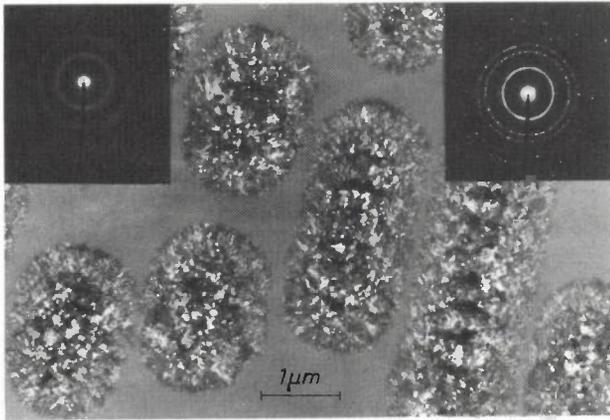


Fig. 10. TEM photograph of crystalline areas written into an amorphous GaSb film of thickness 100 nm. The pattern obtained by local electron diffraction of a heated area (inset top right) is characteristic of a crystalline material; the pattern obtained for an unheated area (inset top left) is characteristic of an amorphous material.

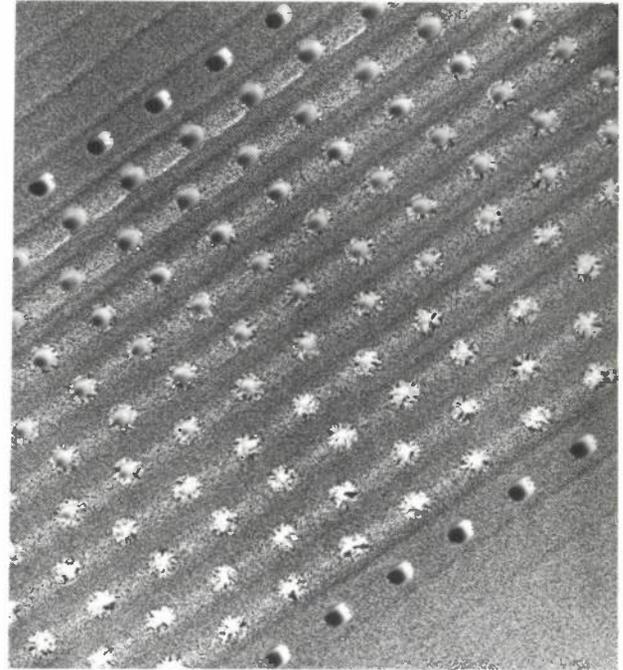


Fig. 11. TEM photograph of tracks in $\text{Te}_{0.80}\text{Se}_{0.10}\text{Sb}_{0.10}$ with written patterns. The top track and the bottom track have not been erased; the others have been exposed to the erasing laser beam for different numbers of times: the second from the top once, the next one twice, and so on. After a single erasure exposure the amorphous areas formed in the writing process are virtually unaffected. After an increasing number of exposures the amorphous areas have become completely crystalline again. These areas can be distinguished from the unwritten areas in this photograph, but when they are read out with a laser the erased areas give the same signal as the unwritten areas. From the intermediate tracks it can be seen that the crystallization in the erasure process starts at the crystalline boundaries of the amorphous areas.

at a maximum power of about 8 mW at the disc, with 80-ns pulses. During writing and read-out the discs rotated at a constant linear velocity of 1.3 m/s.

This recording system was found to be suitable for information storage^[6]. The carrier-to-noise ratio (or CNR) that could be obtained was sufficient: about 55 dB at a repetition rate of 720 kHz and a bandwidth of 10 kHz. Ageing experiments have indicated that the stability of the amorphous film and the written effects are well up to archiving standard. Even after reading out 10^6 times at a power of 0.8 mW the decrease in CNR-value was barely measurable (less than 2 dB).

Erasable optical recording

For erasable optical recording the material of the recording layer must have a crystallization time longer than the cooling time after writing. This is the case with TeSbSe alloys, for example^[5]. Fig. 11 shows a TEM photograph of such an alloy with written patterns that have been exposed to different amounts of erasure. Amorphous areas were produced in the crystalline tracks during the writing. Areas that were sufficiently heated during the erasure have become completely crystalline again. A disadvantage of TeSbSe alloys with adequate stability of the amorphous and crystalline phases at about 70 °C is that the crystallization time in erasure is longer than the pass time of the laser beam. Their crystallization time is longer than 10 μs, while at a typical linear recording velocity of say 1 to 10 m/s the corresponding pass times are only 1 μs and 0.1 μs, respectively.

All the requirements for erasable recording can be met with an InSbTe alloy. Discs with a 100-nm recording layer have been made by evaporation on to a pre-grooved substrate. The discs were tested in an optical recording system with a laser operating at 752 nm and an objective lens with a numerical aperture of 0.5. Since an evaporated InSbTe alloy is amorphous, the tracks were first crystallized during a single revolution of the disc, by c.w. laser radiation at a power of 4.2 mW. To determine the CNR-value a carrier at a frequency of 750 kHz and a pulse duration of 60 ns was written at a constant linear disc velocity of 1.4 m/s. The CNR-value obtained on read-out at a typical bandwidth of 10 kHz depends strongly on the energy at which the information was written into the recording layer; see fig. 12. A value of 52 dB is reached with pulses of energy 0.6 nJ.

We have investigated the effectiveness of the erasing process by erasing the information written at 0.6 nJ at various laser powers during a single revolution of the disc and then determining the residual CNR-values.

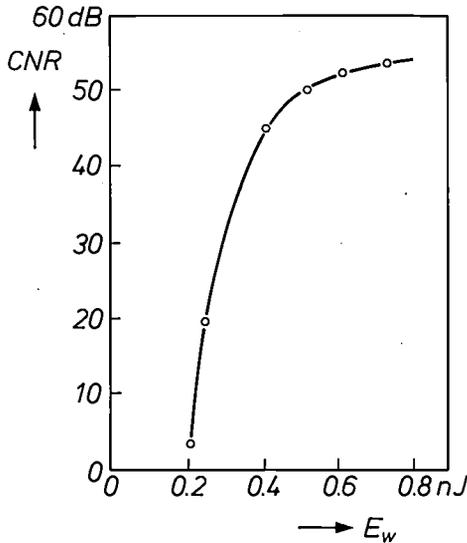


Fig. 12. Carrier-to-noise ratio CNR on read-out with a bandwidth of 10 kHz, as a function of the energy E_w at which the carrier was written into crystalline tracks of an InSbTe alloy (see text). Above a certain threshold energy (about 0.2 nJ) the CNR-value increases rapidly with E_w . At a higher write energy there is some saturation; writing at $E_w = 0.6$ nJ gives a CNR-value of 52 dB.

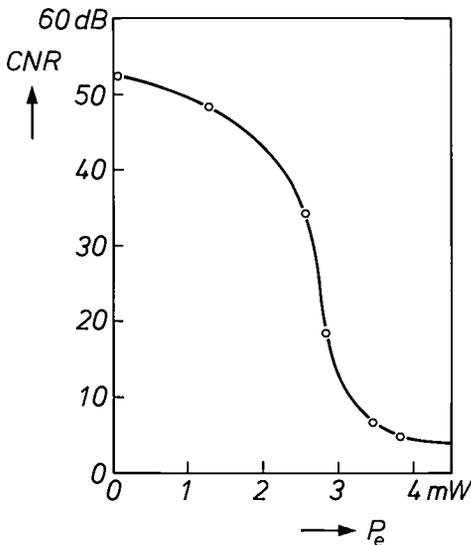


Fig. 13. CNR-value remaining after c.w. erasure during a single revolution of the disc, as a function of the laser power P_e during the erasure process. An increase in this power gives a rapid decrease in the CNR-value: from 52 dB at 0 mW to about 5 dB at 4 mW.

The results are shown in *fig. 13*. After erasure at a laser power of up to 4 mW, the CNR-value falls from 52 dB to about 5 dB. With discs that had no protective coating it was possible to switch between writing and erasure at least 700 times before problems with tracking arose as a result of partial breakaway of the recording layer. The stability of the recorded information was found to be even better than the thermal stability of the plastic substrates; the film can be heated to 140 °C for longer on glass substrates without the information being corrupted.

Prospects

Compared with magneto-optical recording, crystalline/amorphous recording has the advantage that the recording system is simpler and more compatible with existing forms of optical recording. On the other hand, the technology is by no means as far advanced as that for magneto-optical recording, which can offer unlimited writing and erasure. More research will be necessary before higher values for the signal-to-noise ratio can be obtained. The number of times that the system can be switched between writing and erasure will also have to be much larger. There are good prospects for the application of write-once discs with amorphous GaSb for non-erasable optical recording.

Summary. In phase-change optical recording the recording layer of a disc is heated locally by a focused laser beam. Heating crystalline material to just above the melting point and then cooling it rapidly causes it to change from the crystalline phase to the amorphous phase, while heating amorphous material below the melting point has just the opposite effect. The read-out of written effects depends on the difference in optical properties between the two phases. The choice of materials for the recording layer is determined not only by the optical properties, but also by the crystallization behaviour. The stability of amorphous areas and the crystallization rate at higher temperatures are particularly important here. A material with a very high crystallization rate, GaSb, is found to be suitable for non-erasable recording in an amorphous recording layer. For erasable recording in a crystalline recording layer all the requirements can be met with an InSbTe alloy. The first results of recording experiments with these materials look very promising.

Magnetic fields in medical diagnostics: MR and SQUID

O. Dössel, M. H. Kuhn and H. Weiss

The medical application of MR (Magnetic Resonance) has been an important part of the research programme at Philips Forschungslaboratorium Hamburg for a number of years. One of the significant landmarks in this research has been the installation of an MR system with a four-tesla superconducting magnet. MR diagnostics at such high magnetic fields presents a particularly interesting challenge. So too does the measurement of the magnetic field around the human brain, magneto-encephalography, where the flux densities to be detected never exceed 1 picotesla (10^{-12} T). The Hamburg laboratories are developing a magnetometer that measures these extremely weak fields with a SQUID (a Superconducting QUantum Interference Device). The intention is to use this magnetometer to reconstruct images of the neuronal activity in the human brain.

Introduction

In modern medical diagnostics the creation of images of the interior of the human body is of vital importance. X-rays, radioactive isotopes and ultrasound have long been used for such imaging. A recent newcomer to these diagnostic aids is the magnetic field, now used in proton magnetic resonance, or MR (for Magnetic Resonance). However, MR does more than provide a better understanding of the morphology (form and structure) of the body. It also gives information about physiological functions, such as the metabolism. The great advantage of MR is that so far there has been no indication that it is in any way harmful.

MR spectroscopy has been widely used for some ten or twenty years for the investigation of molecular structure in organic chemistry. Recently MR spectroscopy has also been used for the investigation of the chemical functions of human organs. Usually

chemical analysis in medical diagnosis is an *in vitro* technique — a sort of 'test-tube' process. The great advantage of MR spectroscopy is that it is an *in vivo* technique: living tissue can be investigated without the need for surgery.

In MR the atomic nuclei in a small 'slice' of the body are excited by an r.f. magnetic field, which is superimposed on a powerful constant magnetic field. The signals emitted by the excited nuclei are analysed in terms of frequency. The resolution of MR spectroscopy increases with the magnitude of the constant magnetic field. The investigations made at the Philips Hamburg laboratories (Philips Forschungslaboratorium Hamburg) in recent years have therefore mainly concentrated on MR with very powerful magnetic fields. This work has recently culminated in a 'total-body system' that includes a superconducting magnet with a flux density of no less than 4 T (4 teslas or 40 000 gauss); see *fig. 1*. The magnet encloses the patient's entire body.



Fig. 1. The photograph shows the superconducting 4-T magnet — weight 50 tonnes — being lowered through the roof of one of the MR-research buildings at the Philips Hamburg laboratories.

In addition to this work on MR at very high magnetic fields, the Hamburg laboratories are also investigating extremely weak magnetic fields, like those generated by the active areas in the brain^[1]. These only have a flux density of about 10^{-14} T. The device used for measuring these fields is a transducer based on the Josephson effect, and is called SQUID (for Superconducting QUantum Interference Device). Investigations into the activity of the various zones of the brain could be of value in the diagnosis of illnesses like epilepsy and schizophrenia.

Unlike MR, magneto-encephalography (MEG) with SQUIDS is not used for studying the morphology of the human brain, but for localizing the active zones. This is very difficult to do by electroencephalography (EEG), which only measures electrical potentials at the scalp. Another technique, positron emission tomography (PET), only detects the metabolism. Magneto-encephalography will give a direct image based on the currents in the neurons, with a spatial resolution of a few millimetres and a temporal resolution of a millisecond.

In our article we shall first look more closely at magnetic resonance, then at some of the results obtained with our first MR system, which operates at a magnetic field of 2 T. Next we shall show that the chemical-shift effect, which was originally an unwanted effect, can be turned to good use in discriminating between water and fat in MR images. The new magnet system with a flux density of 4 T utilizes the chemical shift in the form of MR spectrometry. Some results of investigations with the 4-T system are also presented. A description is then given of the SQUID system and the biomagnetic measurements and images that can be made with it. Finally, future developments in both areas are considered.

Theoretical principles of magnetic resonance

Magnetic resonance is based on the effect in which the spin axis of the nucleus of a hydrogen atom precesses about the direction of a constant magnetic field. The angular frequency ω of this 'Larmor precession' is proportional to the flux density B_0 of this magnetic field:

$$\omega = \gamma B_0. \quad (1)$$

The quantity γ is the gyromagnetic ratio of the proton. An r.f. magnetic field of flux density B_1 , applied by r.f. coils in a direction perpendicular to the constant field, excites the Larmor precession if its frequency is equal to the precessional frequency $\omega/2\pi$ ^[2].

A specimen in thermal equilibrium in the constant magnetic field has a small magnetization in the direction of the field, since there are more spins aligned with the field than spins aligned in the opposite sense. The application of the r.f. field for a short time makes the magnetization rotate through a particular angle. Because of the Larmor precession of the individual protons, the magnetization vector starts to precess as well. After some time this precession dies away as the system returns to the state of thermal equilibrium. The magnetization in the longitudinal direction of B_0 has a slower relaxation to thermal equilibrium than the transverse magnetization. The time constant T_1 associated with the relaxation of the longitudinal magnetization is therefore larger than the time constant T_2 associated with the relaxation of the transverse magnetization.

The precessing magnetization vector induces a signal in the r.f. coils, and from this r.f. signal values can be derived for the relaxation times T_1 and T_2 and for ρ , the proton density in the specimen. The proton density is proportional to the water content. Spatial information can be added to this signal by means of 'gradient coils', which superimpose a gradient in the direction of one of the spatial coordinates on the constant magnetic field. The resultant slight variation in the field causes a corresponding shift in the frequency of the detected signal. Image-reconstruction methods can be used for translating the signals induced in the r.f. coils into images. The grey-level values in these images can be made dependent on T_1 , T_2 or ρ , or combinations of these quantities^[3].

Results with the 2-T magnet system

In 1983 an MRI Centre (MRI for Magnetic-Resonance Imaging) was set up at Philips Forschungslaboratorium Hamburg. To avoid problems with interfering magnetic fields the centre was housed in two wooden buildings. The centre has two MR systems

that can be used for all the MR applications of interest in medical diagnostics. The first MR images ever made at $B_0 = 2\text{ T}$ were obtained in March 1984^[4]. The quality of the images was far better than we had expected. It had not been entirely clear that the penetration depth of the r.f. magnetic field would be sufficient to excite the hydrogen atoms, since eq. (1) indicated that the frequency of the r.f. field should increase at higher values of B_0 .

Images obtained in close cooperation with medical users at the Hamburg-Eppendorf University Hospital (see fig. 2) also showed that the high flux density of 2 T is not a fundamental limitation. The medical users examined patients in our laboratory before they had their own MR system with a magnetic field of 1.5 T. The four images in fig. 2 all relate to the same cross-section of the brain. They show both the strengths and the weaknesses of MRI: the grey-level values of an MR image can be made proportional to the three tissue parameters T_1 , T_2 and ρ or to a combination of these, as desired. The contrast in the image is determined by the dominant parameter. The parameter ρ depends mainly on the water content; the parameters T_1 and T_2 can often be used to give discrimination between diseased and healthy tissue. The extent to which the three parameters are represented in the image depends largely on the experience of the medical user.

An MR system is a complex assembly of the following units:

- the superconducting magnet for the main field,
- the gradient coils,
- the r.f. coils,
- the spectrometer, which analyses the detected r.f. signals, and
- the computer, which controls the data processing and image reconstruction.

One very important unit is the r.f. coil used for generating and detecting the r.f. field. At a constant main field of 2 T, the frequency of the r.f. field should be 85 MHz.

In the design of the r.f. coil it is important to maximize the sensitivity of the coil and the uniformity of the r.f. magnetic field. The sensitivity affects the signal-to-noise ratio and the uniformity of the r.f. field affects the brightness distribution in the image. It has been found that at high values of the main field the conventional saddle-shaped coils are not satisfactory and that resonators give much better results^[5]. Fig. 3 shows the resonator we have designed for imaging cross-sections of the trunk and the smaller resonator used for imaging cross-sections of the cranium.

When a cross-section of the trunk is produced the power supplied to the r.f. coil is a few kilowatts. Lim-

its are set to this power by the amount of power the patient can absorb without an excessive rise in body temperature. In making localized images of an eye, the spinal column, a breast or a joint, for example,

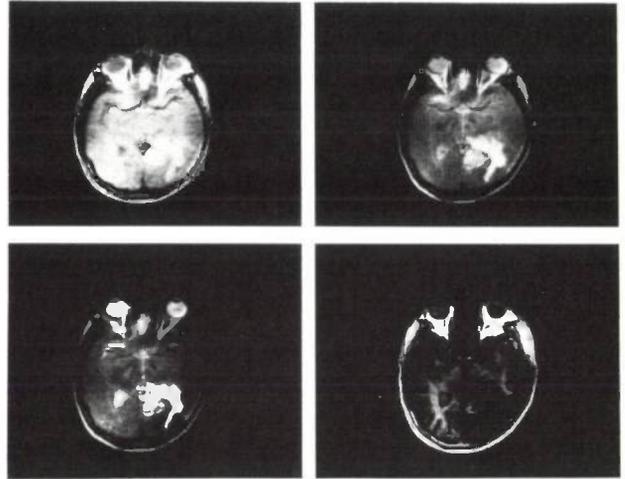


Fig. 2. The first images obtained with the 2-T MR system, made in cooperation with medical users in the Hamburg-Eppendorf University Hospital. The images show cross-sections of the cranium. A tumour surrounded by oedema can be seen. The contrast in the four images is different because the relative contributions from the tissue parameters T_1 , T_2 and ρ in each image are different.

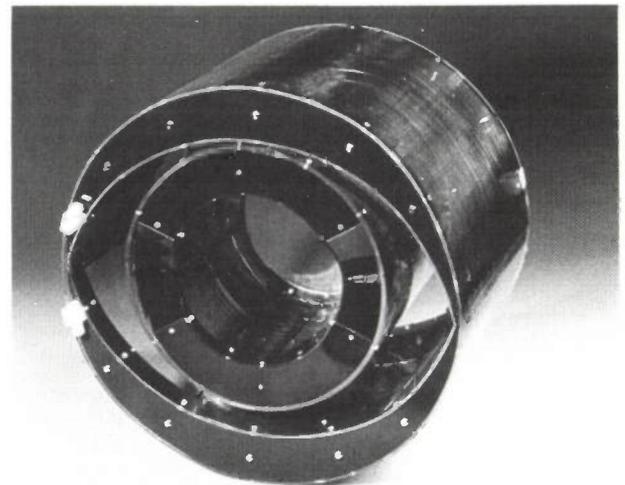


Fig. 3. Resonators for generating and detecting the r.f. field. The outer resonator is used for imaging cross-sections of the trunk, the inner resonator for imaging cross-sections of the cranium.

- [1] S. J. Williamson and L. Kaufman, Magnetic fields of the human brain, *J. Magn. & Magn. Mater.* **15-18**, 1548-1550, 1980.
- [2] P. R. Locher, Proton NMR tomography, *Philips Tech. Rev.* **41**, 73-88, 1983/84.
- [3] M. H. Kuhn and W. Menhardt, Interactive MR image synthesis, *Philips Tech. Rev.* **43**, 95-103, 1987.
- [4] H. Bomsdorf *et al.*, First MR body images at 2 Tesla, *Proc. 3rd Ann. Meeting Soc. Magn. Res. in Medicine*, 61, New York 1984.
- [5] P. Röschmann and C. Wetzel, Die Bedeutung der HF-Spulen für die Kernspintomographie, *Biomed. Tech.* **31**, 178-185, 1986.

the heat dissipated can be greatly reduced by using 'surface coils' [6]. The r.f. field is then relatively non-uniform, see *fig. 4*, which means that only regions just below the surface can be imaged. An advantage of surface coils is that a general image can be obtained first without much power in the coil, see *fig. 5*, and then the image can be 'zoomed in' on a particular organ.

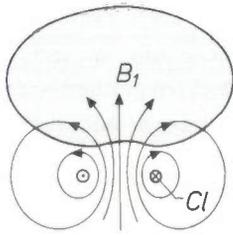


Fig. 4. Non-uniformity in the r.f. field of flux density B_1 when a surface coil C_I is used. The advantage of such a coil is that less heat is dissipated in the patient's body.

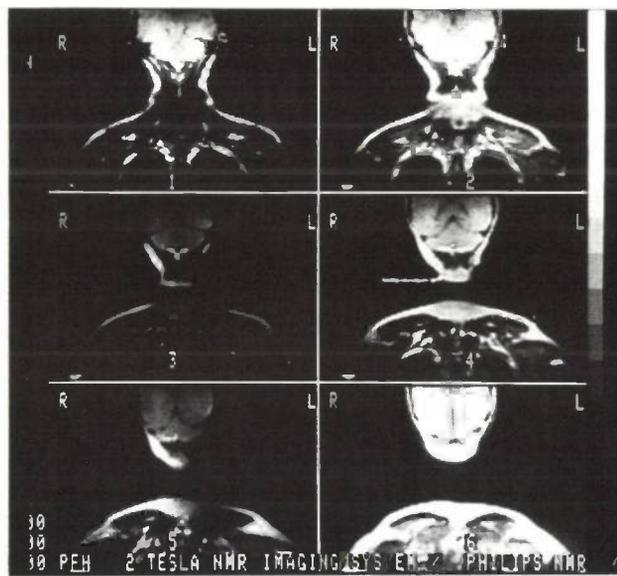


Fig. 5. Images of six different cross-sections of the cranium with part of the trunk, made with a surface coil 35 cm by 40 cm. The dimensions of the field of view are 50 cm by 50 cm, the 'slice thickness' is 7 mm and the pixel size is 2 mm by 2 mm.

Chemical shift and discrimination between water and fat

Chemical shift is the — usually undesirable — slight difference in the precessional frequency of hydrogen nuclei in different molecules, especially those of water and fat. Because of the frequency difference the contributions from the protons in water are displaced slightly in the image with respect to the contri-

butions from the protons in fat, giving a blurred image. It seemed at first that this would degrade the high signal-to-noise ratio obtained with the high main field.

One answer to this dilemma is to make separate images for water and fat. This can be done by the 'CHESS method' (CHESS stands for CHEMical Shift Selection) [7], in which the resonance of the protons in water is suppressed and that of the protons in fat is enhanced, and vice versa. Images resulting from this method are shown in *fig. 6*. The discrimination between water and fat in the image not only solves the problem of the chemical shift but also creates new diagnostic capabilities.

The discrimination between fat and water is a first step towards *in vivo* spectroscopy by MR. In MR spectroscopy the r.f. signal analysed originates from the excited hydrogen nuclei of the tissue in a small region of the body, a region with a volume of a few millilitres. The medical objective here is to provide a better understanding of the chemical processes in the body — something more complicated than making a morphological image.

The spectrum of the signal of the excited hydrogen nuclei shows peaks that belong to different molecule structures. The peaks of medical interest can be resolved if the flux density of the main field is greater than 1 T. Here Philips have already shown with their

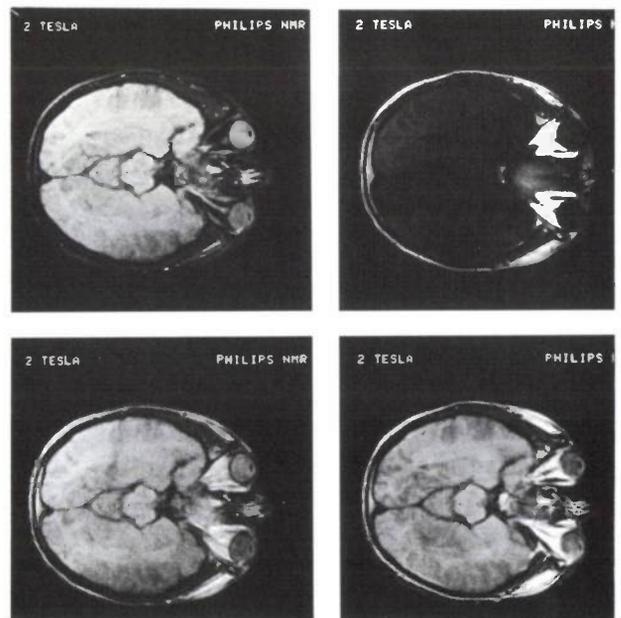


Fig. 6. Discrimination between the contributions from water and fat in a cranial image. Top left: water alone. Top right: fat alone. Bottom left: coincident contributions from water and fat in a particular ratio. Bottom right: the original image, in which the contributions from water and fat do not coincide, so that the picture is blurred.

Gyrosan S15 that they are leaders in combining MR imaging and spectroscopy at 1.5 T^[8]. Combined Philips systems of this type are already in use in clinics in West Germany and other countries. However, the spectroscopic resolution becomes more effective as the flux density increases. This explains the demand for a main field of flux density higher than the two teslas of the MR system already installed in Hamburg. We have therefore developed a scheme for a total-body system with the unprecedentedly high flux density of four teslas.

Development of a 4-T magnet system

In fact it turned out to be possible to build a four-tesla superconducting magnet that would fit round the human body. But it was not entirely clear whether it would also be possible to develop a complete MR system with such a magnet. The crucial question was whether MR images could be obtained at a Larmor frequency as high as 170 MHz, in other words whether an alternating magnetic field at this frequency would penetrate far enough into the body.

The question was answered by combining experiment and theory. The theory indicated that the resonant frequency of the r.f. coils is displaced slightly by currents in conducting parts of the body, for example in muscle tissue^[9]. Measurements of this frequency shift gave a value for the penetration depth of the r.f. field. It appeared that a penetration depth of 10 cm was likely at 170 MHz. It would therefore be possible to obtain images of the cranium, but it would probably not be possible to obtain images of cross-sections of the trunk. Nevertheless it seemed worth while to test these expectations by building an MR system with a flux density of 4 T.

A second problem is the uniformity of the main field. The width of the peaks in the MR spectrum is proportional to the departure from uniformity of this magnetic field. Improving the resolution of the peaks is not very useful if the peaks become broader at the same time, of course. Careful adjustment of the constant main field is therefore an essential condition for reliable spectroscopic analysis. Unfortunately, when the patient is brought inside the magnet the main field becomes less uniform, since the susceptibility of the body is not the same everywhere. Local 'shimming' of the main field is therefore necessary when the patient is placed in the magnet. This is difficult and takes a long time.

For rapid and effective MR spectroscopy with a magnetic field of four teslas we have developed an automatic shimming procedure. In this procedure the Fourier transform of the signal from the excited pro-

tons in a particular slice gives the width of the peaks in the MR spectrum. A computer program matches the currents in the individual coils to minimize the width of the peaks. The result is that the uniformity of the main field in the slice defined is improved by a factor of 10^[10].

Another problem is associated with the gradient coils: currents of up to 500 A have to be switched in less than a millisecond in these coils. At every imaging scan of the body these high currents are responsible for large Lorentz forces on the turns of the coils. We have therefore developed an extremely strong and rigid unit for the three gradient coils for spatial discrimination in the *x*-, *y*- and *z*-directions. The gradient-coil unit is illustrated in *fig. 7*. Another considerable challenge was the development of a spectrometer for signals of frequencies up to 170 MHz. We designed and made a broadband spectrometer for these high frequencies, and this was finally tested in the 2-T system at lower frequencies.

The design and engineering of a 4-T system is in itself a considerable challenge. The superconducting magnet was developed and manufactured by the Oxford Magnet Technology Company to Philips specifications. The magnet is generously dimensioned, and

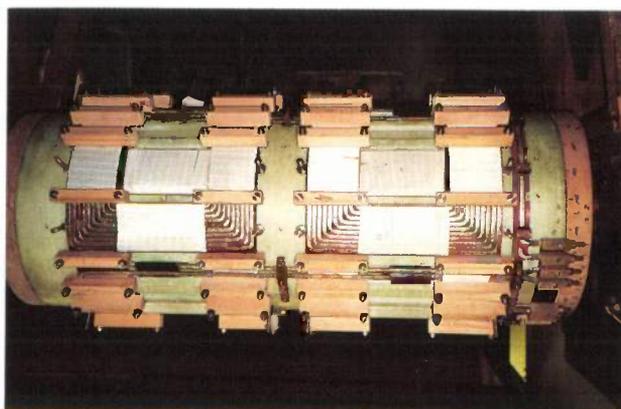


Fig. 7. The unit with the gradient coils for the *x*-, *y*- and *z*-directions of the MR system in which the main field has a flux density of 4 T.

- [6] P. Röschmann and R. Tischler, Surface coil proton MR imaging at 2 T, *Radiology* **161**, 251-255, 1986.
- [7] J. Frahm, A. Haase, W. Heinicke, D. Matthaei, H. Bomsdorf and T. Helzel, Chemical shift selection (CHESS) MR imaging using a 2.0 Tesla whole body magnet, *Radiology* **152**, 79-85, 1985.
- [8] W. Heindel, J. Bunke and W. Steinbrich, Localized ³¹P spectroscopy of the human brain: first clinical results, *Medicamundi* **33**, 50-53, 1988.
- [9] P. Röschmann, Radiofrequency penetration and absorption in the human body: limitations to high field whole body NMR imaging, *Med. Phys.* **14**, 922-931, 1987.
- [10] M. Tochtrop, W. Vollmann, D. Holz and C. Leussler, Automatic shimming of selected volume in patients, *Proc. 6th Ann. Meeting Soc. Magn. Res. in Medicine*, 816, 1987.

provides a uniform magnetic field without automatic shimming. The total weight of the magnet for the main field is 50 tonnes.

Results with the 4-T system

After all the subsystems had been completed, the 4-T system at Philips Forschungslaboratorium Hamburg was brought into operation in October 1987. A photograph of the system is shown in *fig. 8*. MR images of cranial and trunk cross-sections were obtained less than four weeks after the superconducting

the bloodstream with a special fibre-optic sensor and triggering the MR imaging to the bloodstream pulses. The penetration depth, which we had once thought would be the main difficulty at high magnetic fields, is in fact amply sufficient for cranial images. *Fig. 9* shows the first proton images of 5-mm cranial slices.

The 4-T system is mainly intended for easy location of the zone to be analysed by spectroscopic studies. At the same time it can show up abnormalities in the image. A system that can only be used for spectroscopy and offers no imaging facilities would be useless. In the meantime we have obtained the first spectra for

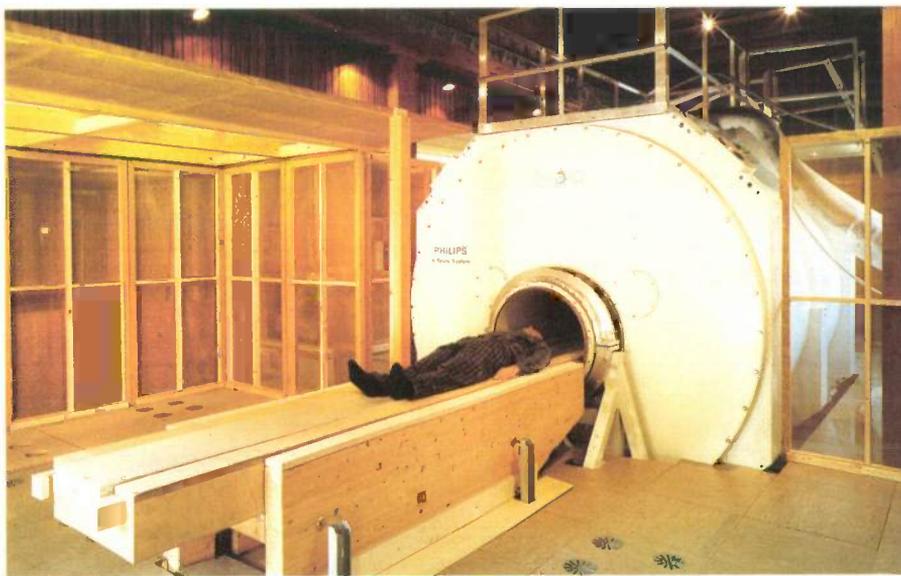


Fig. 8. The MR system with a 4-T superconducting magnet.

magnet for the main field had been delivered^[11]. These images were of very high quality, and any doubts we may have had about imaging with a four-tesla field were instantly removed.

The main objective of the project is not just to produce images but to make spectroscopic analyses with corresponding localization in MR images. We shall make use of precession in the nuclei of phosphorus and carbon atoms as well as in hydrogen-atom nuclei. Because the spectrum peaks are more widely spaced and the signal strengths are higher, the spectral resolution should be better at 4 T than at 1.5 to 2 T.

The first experiments with the 4-T main field confirm our expectations. The images obtained have a high signal-to-noise ratio, but there are some problems with pulse artefacts (image imperfections due to the pulsating H₂O flow in blood vessels). These effects are more noticeable at high magnetic fields. The difficulties were overcome by detecting the pulses in

volume elements with a thickness of 2 cm. These results are better than any previous results in this field. *Fig. 10* shows that the peaks for choline (*Cho*) and phosphocreatine (*PCr*) are clearly separated in the hydrogen spectrum. The first substance plays a part in the transmission of nerve stimuli, the second in the metabolism of muscle tissue.

It will by now be clear that the new 4-T system has added an extra dimension to MR studies — in both diagnosis and technology.

The SQUID system

The magnetic field in the cranium and around it is extremely small. The flux density is in the range 10 fT to 1 pT. Magnetic fields as weak as this can only be detected by using a SQUID^{[1][12]}. This is a detector consisting of a superconducting ring with one or two insulating barriers no more than 1 nm thick. Each

insulating barrier forms a Josephson junction, through which Cooper pairs — paired electrons of opposite spin — can tunnel. The magnetic flux penetrating the ring can be measured by making use of this Josephson effect. The superconducting ring is contained in liquid

helium in a cryostat made of non-magnetic material. The cryostat can be placed close to the patient's head.

As the acronym SQUID indicates, the detector is based on quantum interference effects. These detectors are so sensitive that they can detect changes smaller than $10^{-5} \Phi_0$ in the magnetic flux, where Φ_0 represents the magnetic-flux quantum:

$$\Phi_0 = \frac{h}{2e},$$

where h is Planck's constant and e the charge on an electron. The magnitude of Φ_0 is about 2×10^{-15} Wb or Tm^2 .

The SQUIDs are fabricated in thin-film technology, with niobium nitride separated by magnesium oxide. The properties of a Josephson junction depend very largely on the quality of the extremely thin layer of MgO. The first SQUIDs have recently been completed in Hamburg.

In biomagnetic measurements problems arise from interfering signals due to magnets and electric currents near at hand. We have overcome this difficulty by using filters to screen off fields that are constant in place and time. The equipment is also housed in a wooden building containing no ferromagnetic material. The assembly is mounted on foundations that do not transmit vibration, since movement relative to the earth's magnetic field (5×10^{-5} T) would produce interfering signals. The frame of the installation and the patient table are also made of non-ferromagnetic materials. The frame and the patient table are designed so that the patient's head can be moved into exactly the required position in relation to the detector; see *fig. 11*.

The objectives of the project include not only the measurement of very weak magnetic fields but also the development of a complete system for the diagnosis of diseases of the brain. The biomagnetic measurements made at the Hamburg laboratories will be of great value in adapting the system to the needs of the medical profession. The brain activity of the subjects taking part in the experiments will be measured in response to audible, visible and tactile stimuli.

Biomagnetic measurements and images

Measurements have been made in the SQUID building with a single-channel magnetometer. The hearing of a subject was stimulated by 1000-Hz tones of short duration, and the change in the magnetic field was

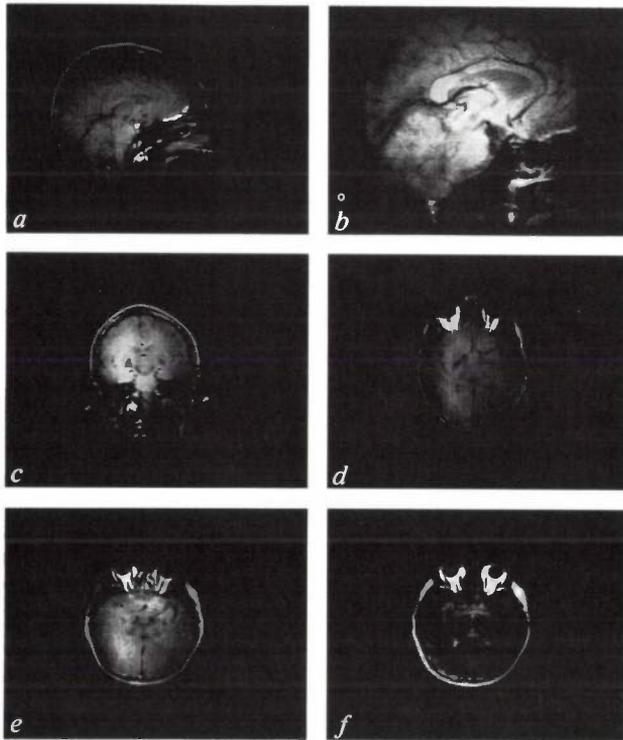


Fig. 9. The first 4-T images of the head. *a*) Sagittal slice, with the imaging triggered to the pulses in the bloodstream. *b*) Enlarged detail of an image with no pulse triggering; the detail image is composed of 512×512 pixels. *c*) Coronal slice with pulse triggering. *d*) Transverse slice with pulse triggering. *e*) Transverse slice made by the spin-echo method and *f*) by the inversion-recovery method [2][3]. (A sagittal plane, a coronal plane and a transverse plane through the head are mutually perpendicular; their directions can be seen from the images shown.)

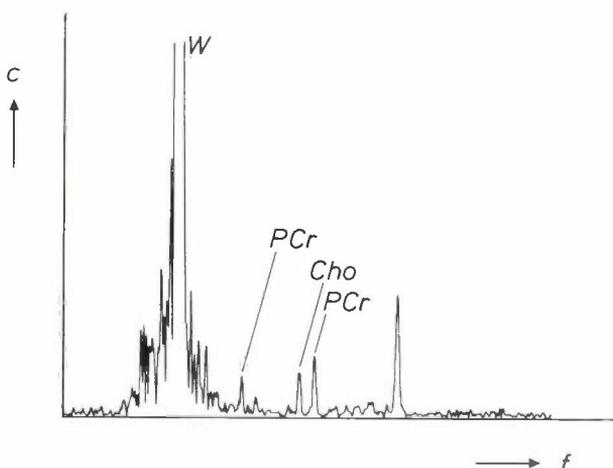


Fig. 10. A spectrum of the signal emitted by excited hydrogen nuclei. *c* relative concentration. *f* relative frequency. *W* water. *PCr* phosphocreatine. *Cho* choline.

[11] H. Bomsdorf *et al.*, MR at 4 Tesla: first results, *Medicamundi* 33, 23-24, 1988.

[12] B. S. Deaver, Jr, *et al.* (eds), Future trends in superconductive electronics, Am. Inst. of Phys., New York 1978.

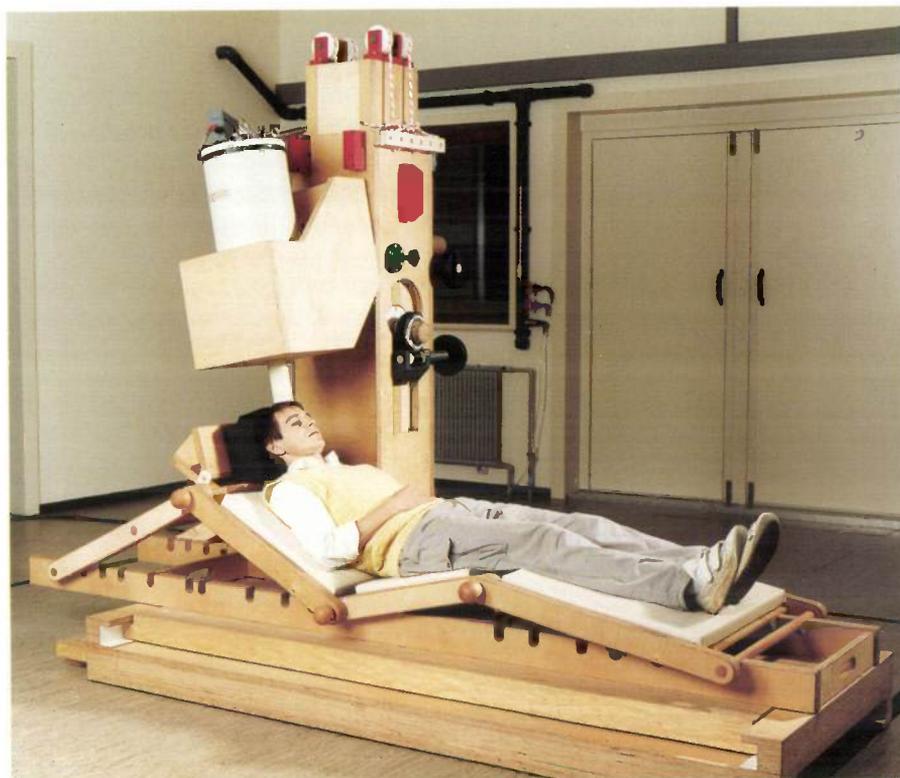


Fig. 11. The SQUID configuration at the Hamburg laboratories. Above the patient's head is the magnetometer, and above that the vessel containing liquid helium. The frame and the patient table are made entirely of non-ferromagnetic material.

measured at three different places near the auditory centre in the cerebral cortex. Results are shown in *fig. 12*. It can be seen that a quiescent period of 100 ms in the signal at *A* is followed by a peak of about 300 fT and then a dip. At *B* there was hardly any change in the magnetic field. At *C* the changes in the magnetic field had the opposite sign to the changes at *A*. These signals characterize an activity in the cortex that corresponds with a neuronal current dipole approximately half-way between the positions *A* and *C*.

The ultimate objective is to reconstruct an image of the current activity in the brain from the changes in the surrounding magnetic field. Unfortunately this 'inverse' problem does not have a unique solution. It is therefore necessary to work from a simplified physiological model. The usual approximation is a linear current dipole in a spherical cranium filled with a conducting medium to complete the circuit for the currents. This model is only a very rough approximation to the real thing. We therefore tried to find a reconstruction algorithm that would give a better approximation.

The result was an image-reconstruction algorithm based on a combination of the Fourier transform of

Maxwell's equations and Whittaker and Shannon's sampling theorem^[13]. The space for the simulation is divided into two zones: the measurement zone and the forbidden zone. The measurement zone is accessible for the measurement of magnetic fields, and the forbidden zone is the zone that contains the unknown current distribution. A plane in which the current distribution is reconstructed can be moved through the forbidden zone, thus giving the current distribution for this zone. Simulations with different current distributions have shown that our reconstruction algorithm is fairly insensitive to interfering magnetic fields.

Future developments

The spectroscopic results obtained in the meantime by a team of physicists, biochemists and doctors have been so encouraging that we are installing 'clones' of our system at clinical centres for further investigations. Since such ultra-high-field MR systems are very expensive the market is limited.

However, the discovery of ceramic materials that are superconductors at higher temperatures could have an influence on the future development of MR

magnets. At the moment the magnet for the main field of an MR system accounts for about half the total cost. If MR magnets could be built that were cooled with liquid nitrogen instead of liquid helium, MR

diagnostics could become less expensive. This would be particularly significant for the more advanced systems.

It has already been shown that these new ceramic materials can be used in SQUIDs. The cost aspect and the fact that a cryostat cooled by liquid nitrogen can be placed nearer to the patient's head will contribute to the widespread future introduction of SQUID systems based on the new materials. One physical limitation, however, is that the noise increases with temperature. This probably rules out the use of SQUID magnetometers at room temperature, but there is hope that they can be used with liquid-nitrogen cooling.

In conclusion, the development of MR imaging and spectroscopy systems continues to go from strength to strength, and the prospects for SQUID imaging systems are most promising.

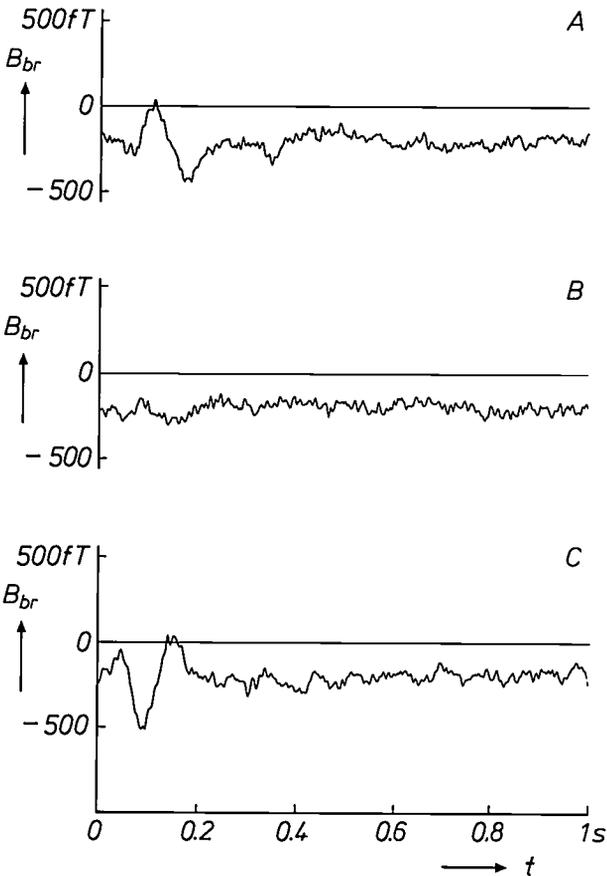


Fig. 12. The flux density B_{br} of the magnetic field near the auditory centre in the cerebral cortex as a function of time t . The field was measured at three different locations: A, B and C.

Summary. A magnetic-resonance (MR) system with a two-tesla superconducting magnet was installed at the Philips laboratories in Hamburg in 1983. Good images of cross-sections of the human body were produced at this high field, but there were difficulties due to the chemical-shift effect, the slight difference in the precessional frequency of hydrogen nuclei in different molecules. This effect gave a blurred image because the images due to water and fat did not coincide. The difficulty was resolved by suppressing one of these images. This was a first step in the development of MR spectrometry, an analytical method that gives information about the chemical processes in a small region of the human body. As the main field increases the resolution of MR spectrometry improves. An MR system with the extremely high flux density of four teslas was therefore developed and brought into operation in 1987. Another development from the Hamburg laboratories is a SQUID magnetometer (SQUID stands for Superconducting QUantum Interference Device), which will be used for detecting the extremely weak magnetic field generated by the currents in the neurons of the brain. A SQUID consists of a superconducting ring with two Josephson junctions, each formed by an insulating film only 1 nm thick. The SQUID will be used in reconstructing an image of neuronal activity from measurements of the field around the brain. A special new algorithm has been developed for use in the reconstruction.

[13] W. J. Dallas, Fourier space solution to the magnetostatic imaging problem, *Appl. Opt.* 24, 4543-4546, 1985; W. Kullmann and W. J. Dallas, Fourier imaging of electrical currents in the human brain from their magnetic fields, *IEEE Trans. BME-34*, 837-842, 1987.

The application of semiconductor superlattices to short-wavelength lasers

P. Blood, C. T. Foxon and E. D. Fletcher

Within the Philips research organization the Laboratories at Redhill, England, are responsible for work on the growth of multilayered structures of III-V semiconductors such as GaAs and $Al_xGa_{1-x}As$ by molecular beam epitaxy. The Laboratories have enjoyed a leading position in this field for many years. Using advanced characterization and analysis methods it is now possible to grow structures with individual layers as thin as a few monolayers with atomically abrupt interfaces. 'Superlattices' consisting of alternate ultra-thin layers of two dissimilar semiconductors with different band gaps are especially interesting. In this article the authors describe an investigation of these structures for application to short-wavelength semiconductor lasers as used in optical-recording systems.

Introduction

Semiconductor lasers are important components in a variety of professional and consumer electronic systems. Lasers based on the $In_xGa_{1-x}As_yP_{1-y}$ materials system and operating at a wavelength of 1.3 to 1.5 μm are used as the light sources in fibre-optic communications systems; shorter-wavelength devices based on the $Al_xGa_{1-x}As$ materials system and operating at wavelengths below 880 nm are used in optical-disc information-storage systems for both the writing and reading operations. In the consumer area these systems are best known in the form of the Compact Disc player, though lasers have other applications such as bar-code readers and laser printers, for example. It is also envisaged that lasers may be used for optical interconnections between silicon-based integrated circuits. In all these applications the semiconductor laser is an attractive component because it is a compact, light-weight, low-voltage, high-brightness, coherent light source.

The detailed specification of a laser depends upon the application, though in most cases a low operating current is a desirable attribute, if only to reduce the power consumption and heat dissipation. This is determined by the threshold current above which laser

operation occurs. A high power output (> 50 mW) is required for operations such as writing information on discs, though devices with a low power output (≈ 5 mW) are quite satisfactory for reading stored information from the disc. Communications systems call for long-wavelength devices operating at the wavelength where the loss and dispersion in the fibre are low, whereas in optical-disc information systems the need is for short-wavelength lasers in order to obtain high storage densities^[1]. Considerable versatility is therefore required to produce types of lasers which are suitable for all these applications, though at a fundamental level there are two requirements which are of primary importance: an appropriate emission wavelength and an acceptably low threshold current. At a development level a high process yield and a long operating lifetime are crucial in the commercial success of all these components.

The fundamental properties of a laser are determined by the properties of the semiconducting materials used in the device, and by the design of the multilayer structure in which they are grown and subsequently processed. Both the intrinsic properties of the materials, primarily their electronic structure and band gap, and the extrinsic properties, such as internal luminescence efficiency, are important in deter-

Dr P. Blood, Dr C. T. Foxon and Dr E. D. Fletcher are with Philips Research Laboratories, Redhill, Surrey, England.

mining the device characteristics. In this article we are concerned with the fundamental properties of semiconductor lasers and with an exploration of the benefits which arise from the use of superlattice structures consisting of a stack of alternate ultra-thin layers of two dissimilar semiconductors with different band gaps. The specific devices we have studied have been fabricated in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ materials system for operation at wavelengths below 880 nm.

Thin-layer structures in lasers

Two kinds of structure are of interest for the devices described in this article: the quantum well, where the carriers are confined within thin layers of GaAs by barrier layers of $\text{Al}_x\text{Ga}_{1-x}\text{As}$, and superlattices comprising alternate thin layers of GaAs and AlAs, sufficiently thin that the carriers are not necessarily localized in one material.

In a quantum well, the charge carriers (electrons in the conduction band and holes in the valence band) are confined in the narrow-gap GaAs 'well' and their momentum perpendicular to the plane of layers is quantized. The associated discrete energy levels depend upon the thickness of the quantum well. These effects occur in layers thinner than about 20 nm. The use of quantum-well structures as the active region of a semiconductor injection laser offers a number of potential advantages over bulk materials: it is possible to adjust the laser wavelength by changing the thickness of the quantum well, the threshold current should be lower, and the temperature dependence of the threshold current should be weaker. Although these advantages are not always realized in practice, quantum wells have certainly proved beneficial in the fabrication of devices which operate at very low threshold current or with high output power, and they have important application in short-wavelength devices [2].

The advantages listed above all derive from the unique electronic properties of a quantum-well structure. However, it is known that by growing a multilayer structure using molecular beam epitaxy (MBE) the properties of the materials in the structure are also improved. For example, it has been shown that the rate of non-radiative recombination of electrons and holes at the interfaces in GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ double heterostructures can be reduced significantly by incorporating a thin multiple-layer structure just below the first interface [3]. In cooperation with K. Woodbridge of these laboratories we have found that the use of such a prelayer in a conventional double-heterostructure laser can reduce the threshold current, the degree of improvement being dependent upon the conditions in the growth system. Multilayer structures also produce an improvement in surface morphology of the

crystal. This is important since it has been demonstrated that thickness variations along a quantum well can have an adverse effect on the laser threshold current [4]. It is therefore possible that the growth of a laser structure in the form of many thin layers is itself beneficial, irrespective of any benefits which may be conferred by the electronic properties of the multilayer structure.

Considerations of this kind have led to the use of all-binary $(\text{AlAs})_n(\text{GaAs})_m$ multilayer structures (n and m are the number of monolayers of each component) to replace the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy regions of a conventional semiconductor laser, and to replace the alloy barrier regions of quantum-well lasers. The simplest approach to these structures is to regard the multilayer as equivalent to an alloy with an Al content equal to the average content \bar{x} of the superlattice, $\bar{x} = n(m+n)^{-1}$, and this is a reasonable approximation as far as the chemical properties are concerned. Fig. 1 shows the Al depth profile of a laser structure consisting of a GaAs quantum well and barrier regions made up of an $(\text{AlAs})_3(\text{GaAs})_9$ superlattice having $\bar{x} = 0.25$, instead of a random-alloy $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$.

In the laser it is important that the injection current can flow across the multilayer structure from the barrier regions to the active region of the device, so the wide-band gap AlAs layers must be sufficiently thin that carriers can tunnel through them. In such a coupled multilayer structure, or electronic superlattice, with binary layers only a few monolayers thick, the electronic properties are not the same as in the average alloy but depend upon the actual values of n and m , not simply their ratio. For example the effective band gap of the superlattice is usually less than that of a random alloy with the same average composition. These effects offer the additional possibility of using the structures to produce superlattice regions with special electronic properties which cannot be obtained with random alloy materials.

The interest in using superlattices in semiconductor lasers is therefore generated by two factors: improvements in extrinsic material properties, such as morphology, due to the multilayer structure, and tailoring the intrinsic electronic properties by adjusting the

[1] G. E. Thomas, Future trends in optical recording, Philips Tech. Rev. 44, 51-57, 1988.

[2] P. Blood, Reappraisal of GaAs-AlGaAs quantum well lasers, Int. Symp. on Technologies for optoelectronics, Cannes 1987, Proc. SPIE 861, 34-41, 1988.

[3] P. Dawson and K. Woodbridge, Effects of prelayers on minority-carrier lifetime in GaAs/AlGaAs double heterostructures grown by molecular beam epitaxy, Appl. Phys. Lett. 45, 1227-1229, 1984.

[4] P. Blood, S. Colak and A. I. Kucharska, Influence of broadening and high-injection effects on GaAs-AlGaAs quantum well lasers, IEEE J. QE-24, 1593-1604, 1988.

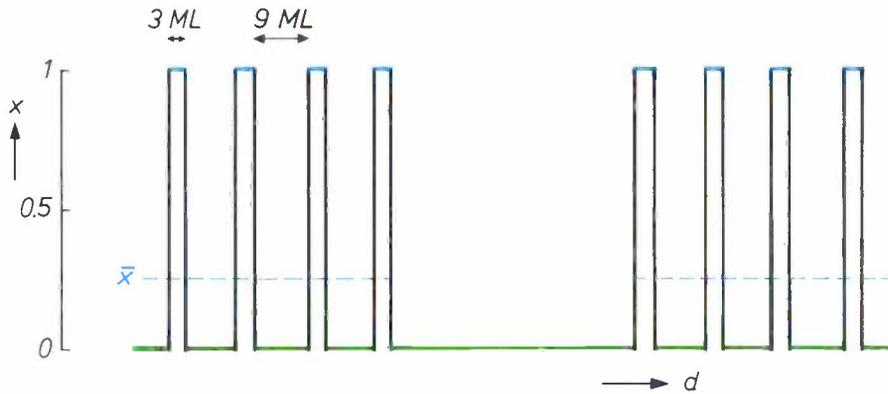


Fig. 1. Composition profile of a multilayer laser structure in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ materials system. The Al content x of the various layers is plotted as a function of the distance d from the substrate surface. The structure comprises a GaAs quantum well and $(\text{AlAs})_3(\text{GaAs})_9$ superlattice barriers formed by alternating three monolayers (ML) of AlAs and nine monolayers of GaAs; the average Al content \bar{x} of the barriers is 0.25.

thicknesses of both AlAs and GaAs layers. The principal practical problem to surmount is the controlled growth of a structure containing up to several thousand layers, each only a few monolayers thick. Molecular beam epitaxy (MBE) has proved well-suited to this task. *Fig. 2* is a micrograph, obtained by trans-

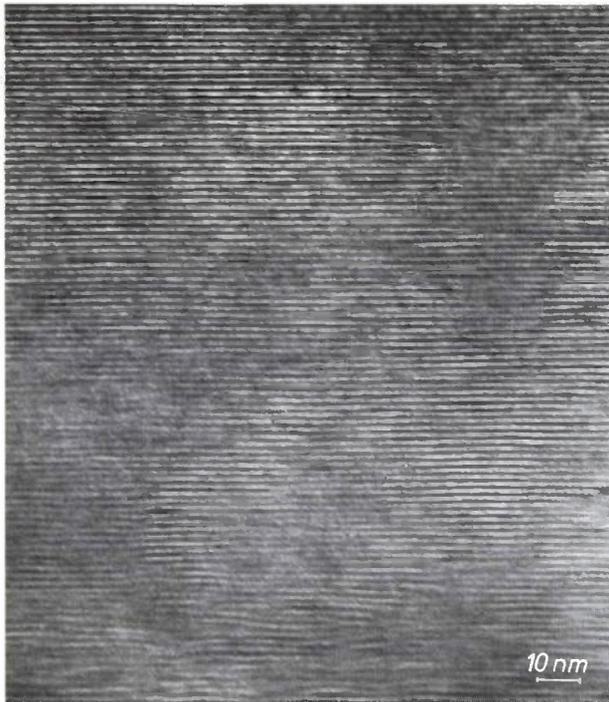


Fig. 2. Cross-sectional (002) dark-field micrograph, obtained by transmission electron microscopy, of a superlattice grown by molecular beam epitaxy. The superlattice comprises 100 periods of alternate layers of AlAs and GaAs, each three monolayers thick. The micrograph was taken by J. P. Gowers at PRL.

mission electron microscopy (TEM), of a superlattice structure grown by MBE and comprising alternate layers of AlAs and GaAs each as thin as 0.8 nm, corresponding to three monolayers of material.

In this article we describe the growth and device performance of quantum-well lasers embodying all-binary superlattice barrier and cladding regions, and of conventional semiconductor lasers incorporating a superlattice active region. These devices have been grown by MBE by D. Hilton and processed by P. J. Hulyer, and all incorporate superlattices with binary layers as thin as three monolayers. The objective of this work is to understand the influence of the superlattices on the performance of the devices.

MBE growth of quantum wells and superlattices

Molecular beam epitaxy (MBE) is a refined form of ultra-high-vacuum (UHV) evaporation. In MBE individual atomic or molecular beams are formed from thermal evaporation sources. These beams impinge on a heated single-crystal substrate to form an epitaxial deposit, with the arrival of unwanted impurities being minimized by the UHV environment. Several excellent review articles have discussed in detail the fundamental factors controlling growth and dopant incorporation in MBE [5], and the growth and properties of superlattices [6]. Recent progress in this field has been summarized elsewhere [7]. Growth rates in MBE are typically one monolayer per second (1 ML/s), corresponding to about one micron per hour and, because molecular beams can be started or stopped rapidly (e.g. within 0.3 s) using a simple mechanical shutter, atomically sharp interfaces can be produced.

Here we concentrate on the practical issues required in growing the type of superlattice structure used in the two types of device discussed in this article. The growth of multilayer structures of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ by MBE has been discussed recently in more general terms in a previous article in this journal [8] and the particular procedure used for this work has been summarized elsewhere [9].

The multilayer structures are grown on the (001) surface of a silicon-doped GaAs substrate with a low dislocation density. The substrate is first chemically cleaned, etched, and mounted using liquid indium on to a molybdenum platen. This is heated to about 400 °C in a subsidiary preparation chamber to remove water vapour, and the native oxides are then removed from the substrate by heating in an excess flux of As at about 600 °C in the MBE growth system.

The structures to be discussed below consist of layers of the binary compounds GaAs and AlAs. Growth rates for these compounds are determined by the net incorporation of the group-III elements Ga and Al, providing an excess As flux is supplied. It has previously been shown that to obtain low-threshold-current lasers it is necessary to grow films at a substrate temperature of about 680 °C [10]. Since the vapour pressure of Ga over GaAs is significant at this temperature, Ga is lost by re-evaporation before being incorporated. As a consequence, for growth of GaAs the net incorporation rate of Ga is less than the arrival rate, reducing the growth rate by approximately 0.2 ML/s [11].

Growth rates for GaAs and AlAs are set using *in-situ* reflection high-energy electron diffraction (RHEED) oscillation measurements on a monitor sample [12]. To achieve a net growth rate of 1 ML/s for GaAs at 680 °C the Ga flux is adjusted to give a measured growth rate of 1.2 ML/s at 600 °C, to allow for Ga re-evaporation; the AlAs growth rate is set to 1 ML/s. The As flux is adjusted to the minimum required to avoid formation of a metal-stable surface at high substrate temperatures. Films are doped n and p type using beams of Si and Be respectively [13], both of which have a unity sticking coefficient at 680 °C.

For all laser structures grown by MBE the quality of the AlAs has a major influence on the threshold current of the devices [14]. This in turn is mainly determined by the incorporation rate of oxygen from the MBE environment [15]. To avoid problems from residual oxygen incorporation it is essential to have a leak-free MBE system, and to minimize the clean-up time for the equipment an adequate bake-out schedule is required. In practice the oxygen incorporation rate can be assessed by measuring the minority-carrier life-

time at room temperature in beryllium-doped multiple quantum-well (MQW) test structures. A long lifetime, indicating low oxygen incorporation, correlates well with a low threshold current for device structures grown in succeeding runs.

The depth profiles of the average composition through two types of superlattice laser are shown in fig. 3. Diagram (a) represents a quantum-well laser containing two wells, each 5.8 nm thick, with $(\text{AlAs})_3(\text{GaAs})_9$ barriers and outer cladding regions each comprising 300 periods of $(\text{AlAs})_6(\text{GaAs})_6$; there is an extra step in the profile of this waveguide made up of a region with the intermediate composition $(\text{AlAs})_3(\text{GaAs})_6$. Diagram (b) represents an all-binary superlattice version of a conventional 'separate-confinement heterostructure' (SCH) laser in which the active region itself comprises a superlattice of $(\text{AlAs})_2(\text{GaAs})_8$ with a thickness of 50 nm. The barrier regions for confining the carriers in the active region are superlattices of $(\text{AlAs})_3(\text{GaAs})_6$ and the light-confining cladding regions are synthesized with superlattices of $(\text{AlAs})_6(\text{GaAs})_6$. Broad-area devices having different cavity lengths were made from these structures and the threshold current was determined by pulsed measurements.

- [6] A. Y. Cho and J. R. Arthur, Molecular beam epitaxy, *Prog. Solid State Chem.* **10**, 157-191, 1975;
B. A. Joyce, Molecular beam epitaxy, *Rep. Prog. Phys.* **48**, 1637-1697, 1985.
- [6] A. C. Gossard, Molecular beam epitaxy of superlattices in thin films, in: *Treatise on Materials Science and Technology*, Vol. 24, K. N. Tu and R. Rosenberg (eds), Academic Press, New York 1982, pp. 13-65.
- [7] Proc. 4th Int. Conf. on Molecular beam epitaxy, York 1986, C. T. Foxon and J. J. Harris (eds), *J. Cryst. Growth* **81**, 1-580, 1987.
- [8] B. A. Joyce and C. T. Foxon, Molecular beam epitaxy of multilayer structures with GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$, *Philips Tech. Rev.* **43**, 143-153, 1987.
- [9] C. T. Foxon and J. J. Harris, The growth of high purity III-V structures by molecular beam epitaxy, *Philips J. Res.* **41**, 313-324, 1986.
- [10] W. T. Tsang, F. K. Reinhart and J. A. Ditzenberger, The effect of substrate temperature on the current threshold of GaAs- $\text{Al}_x\text{Ga}_{1-x}\text{As}$ double-heterostructure lasers grown by molecular beam epitaxy, *Appl. Phys. Lett.* **36**, 118-120, 1980.
- [11] C. T. Foxon, Dynamic aspects of growth by molecular beam epitaxy, in: *Heterojunctions and semiconductor superlattices*, G. Allan *et al.* (eds), Springer, Berlin 1986, pp. 216-255.
- [12] J. H. Neave, B. A. Joyce, P. J. Dobson and N. Norton, Dynamics of film growth of GaAs by MBE from RHEED observations, *Appl. Phys. A* **31**, 1-8, 1983.
- [13] A. Y. Cho and I. Hayashi, Epitaxy of silicon doped gallium arsenide by molecular beam method, *Metall. Trans.* **2**, 777-780, 1971;
M. Ilegems, Beryllium doping and diffusion in molecular-beam epitaxy of GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$, *J. Appl. Phys.* **48**, 1278-1287, 1977.
- [14] W. T. Tsang, The influence of bulk nonradiative recombination in the wide band-gap regions of molecular beam epitaxially grown GaAs- $\text{Al}_x\text{Ga}_{1-x}\text{As}$ DH lasers, *Appl. Phys. Lett.* **33**, 245-248, 1978.
- [15] C. T. Foxon, J. B. Clegg, K. Woodbridge, D. Hilton, P. Dawson and P. Blood, The effect of the oxygen concentration on the electrical and optical properties of AlGaAs films grown by MBE (summary abstract), *J. Vac. Sci. & Technol. B* **3**, 703, 1985.

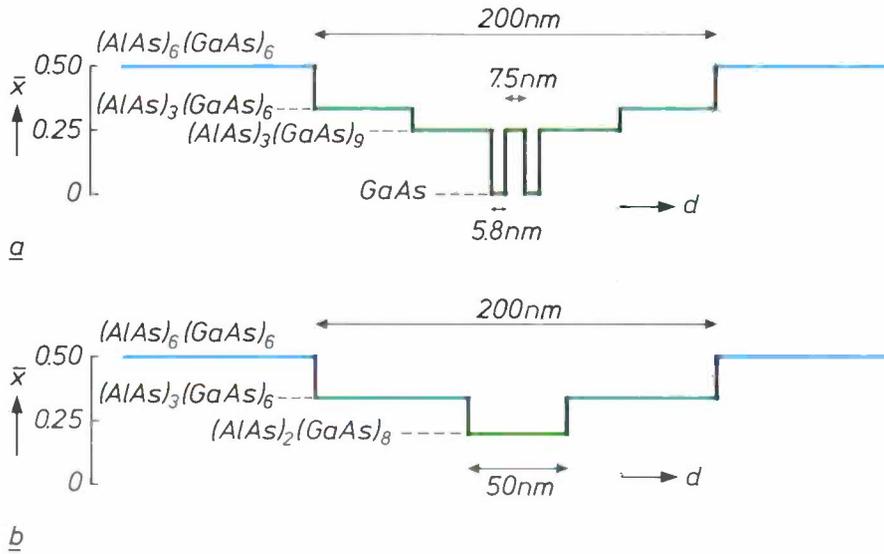


Fig. 3. Profiles of the average Al content \bar{x} through two superlattice laser structures in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ system. *a*) Quantum-well laser containing an active region with two GaAs quantum wells each 5.8 nm thick, barrier regions of $(\text{AlAs})_3(\text{GaAs})_9$ superlattices ($\bar{x} = 0.25$), outer cladding regions of $(\text{AlAs})_6(\text{GaAs})_6$ superlattices ($\bar{x} = 0.50$) and an extra step with the intermediate composition $(\text{AlAs})_3(\text{GaAs})_6$ with $\bar{x} = 0.33$. *b*) Separate-confinement laser with a 50 nm thick active region of an $(\text{AlAs})_2(\text{GaAs})_8$ superlattice ($\bar{x} = 0.20$), 75 nm thick carrier-confining barriers of $(\text{AlAs})_3(\text{GaAs})_6$ and light-confining cladding regions of $(\text{AlAs})_6(\text{GaAs})_6$.

Quantum-well lasers

The quantum-well lasers with superlattice barriers operate at a wavelength of about 830 nm with a threshold current density of 360 A/cm² for a two-well device, which is about 0.7 of the threshold current of a similar alloy device incorporating a superlattice pre-layer, grown consecutively in the same machine. In these devices the wavelength is controlled by the effective thickness of the quantum wells (fig. 3*a*).

An examination of these structures using electrochemical etching and photovoltage spectroscopy [16] shows that the effective band gap of the barrier regions with the average Al content $\bar{x} = 0.25$ is somewhat less than that of an alloy with $x = 0.25$, and corresponds to an alloy with $x \approx 0.18$ -0.20. This is qualitatively in agreement with calculations made by G. Duggan (PRL) on the electronic structure of these superlattices. This reduction in effective band gap of the barrier may cause a small change in the gain-current relation of the quantum-well laser, and may modify any unwanted carrier leakage currents due to recombination in the barrier regions.

To explore these matters further we have measured the threshold current of lasers with different cavity lengths from 150 μm to 1000 μm . Using a nonlinear relation between gain and current which takes account of gain saturation [17], the dependence of the threshold current density J_{th} on the cavity length L_c is given by:

$$\ln J_{\text{th}} = \left(\frac{\ln(R^{-1})}{g_t N \Gamma_w} \right) \frac{1}{L_c} + \frac{\alpha_i}{g_t N \Gamma_w} + \ln(J_t N / \eta_i) - 1, \quad (1)$$

where g_t and J_t are parameters which define the gain-current relation for a single well, R is the fraction of radiation reflected at the facets, N is the number of quantum wells, Γ_w is the optical-confinement factor per well, α_i is the optical-scattering loss per unit length and η_i is the internal luminescence efficiency. According to this relation a plot of $\ln J_{\text{th}}$ versus L_c^{-1} should be linear, with a single-well laser having a line with twice the slope of a double-well laser. These features are apparent in the data plotted in fig. 4. From these slopes we deduce $g_t = 1900 \text{ cm}^{-1}$, which is close to the value of 1730 cm^{-1} measured for an alloy-barrier device.

The intercepts of these plots are more difficult to interpret, though we can eliminate J_t by noting that the difference between the intercepts (I_2, I_1) of an $N = 2$ and an $N = 1$ laser is:

$$I_2 - I_1 = - \frac{\alpha_i}{2g_t \Gamma_w} + \ln(\eta_1 / \eta_2) + \ln 2, \quad (2)$$

where η_1 and η_2 are the internal luminescence efficiencies of the devices. In the presence of barrier leakage currents we cannot assume that η_1 and η_2 are the same, though at low temperatures where the populations of carriers in the barriers become negligibly small we do expect that $\eta_1 = \eta_2$. In this case

$\ln(\eta_1/\eta_2) = 0$ and we can determine α_i from equation (2). From measurements of J_{th} as a function of cavity length at the reduced temperature of 160 K, plots similar to those in fig. 4 give a value for $I_2 - I_1$ from which we find that $\alpha_i \leq 3 \text{ cm}^{-1}$. This is the value expected for free-carrier scattering alone and suggests there is no scattering loss due to the various interfaces in the structure.

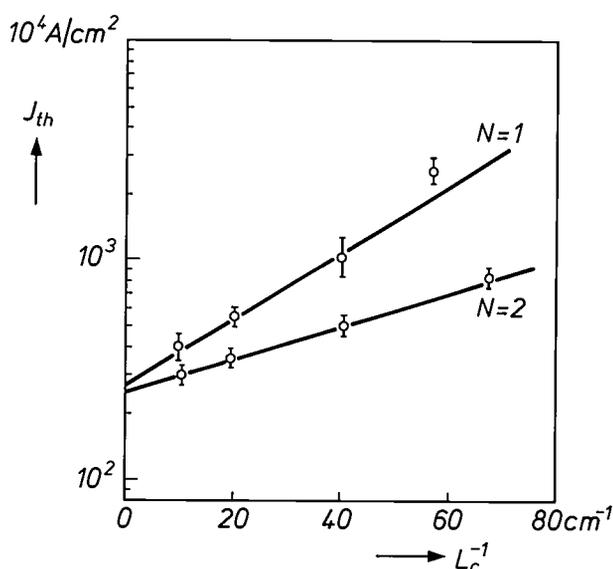


Fig. 4. Log-linear plot of the threshold current density J_{th} of superlattice quantum-well lasers with one well ($N = 1$) and with two wells ($N = 2$) measured at 300 K and plotted as a function of L_c^{-1} , the reciprocal of the cavity length. Both lasers give a straight line and the slope for $N = 1$ is twice that for $N = 2$, in agreement with equation (1).

At higher temperatures, $I_2 - I_1$ gets smaller which indicates that η_1 is becoming smaller than η_2 (equation 2) and this is consistent with a carrier leakage process in the superlattice barrier regions. This leakage process may represent a significant fraction of the threshold current at room temperature, especially in the single-well device, and we believe that it is related to the observation that the superlattice barrier has a low effective band gap compared with the alloy. It should be possible to modify the design of the barrier to reduce this leakage current.

Separate-confinement all-superlattice laser

The quantum-well laser offers an attractive means of shortening the laser wavelength by reducing the width of the well while at the same time retaining a

GaAs recombination region. The long-established method of shortening the wavelength in a conventional double-heterostructure laser is to use wider-band-gap $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ($x > 0$) in an active region 50 nm to 200 nm thick; in this way emission at about 780 nm can be obtained with $x \approx 0.13$. One difficulty sometimes encountered with this approach is the problem of growing an $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy with a good internal luminescence efficiency, so we have made a conventional laser with all the alloy regions synthesized throughout by all-binary (AlAs)(GaAs) superlattices.

The structure is shown in fig. 3b and incorporates a superlattice active region 50 nm thick with an average Al content of 0.20. The device operates at a wavelength of 786 nm and we estimate that these initial structures are of similar quality to the materials used in present commercial devices. In this structure the light is generated in a superlattice, though we note that the average Al content \bar{x} is somewhat greater than the Al content x of the random-alloy material required for emission at the same wavelength. Although research on these devices is at an early stage, their threshold current is sufficiently low to encourage further work on the growth and physics of these superlattice lasers.

These devices show that it is possible to make lasers using all-binary superlattices throughout, instead of alloys, in which the optical gain is also generated within a superlattice active region. As with the superlattice quantum-well lasers, work is now directed toward optimizing the performance of the devices by using more precise theoretical calculations of the electronic band structures of superlattices to improve the design of the superlattice regions in the devices.

A general account of the calculation of the electronic properties of complex material systems, including (AlAs)(GaAs) superlattices, is given in the following article (pp. 276-286).

Summary. This research shows that short-period superlattice structures can be used to advantage in a variety of short-wavelength semiconductor lasers. Molecular beam epitaxy has been used to grow laser structures in which the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy regions have been replaced by all-binary short-period (AlAs) $_n$ (GaAs) $_m$ superlattices embodying binary layers as thin as three monolayers. Quantum-well lasers with superlattice barriers and GaAs wells have a lower threshold current than equivalent structures using alloys. An analysis of the performance of devices with different cavity lengths shows that the optical-scattering losses in these structures are indeed very low. There is evidence for a leakage current associated with recombination in the barrier regions though it should be possible to reduce this by modifying the design of the superlattice barrier. Double-heterostructure lasers have been made with an all-binary superlattice active region which operates at 786 nm and with encouraging values of threshold current.

[16] P. Blood, Capacitance-voltage profiling and the characterisation of III-V semiconductors using electrolyte barriers, *Semicond. Sci. Technol.* 1, 7-27, 1986.

[17] P. W. A. Mc Ilroy, A. Kurobe and Y. Uematsu, Analysis and application of theoretical gain curves to the design of multi-quantum-well lasers, *IEEE J. QE-21*, 1958-1963, 1985.



Briarcliff Manor

Philips Laboratories, North American Philips Corporation, Briarcliff Manor, NY, U.S.A.

High-power integrated electronics, television systems (HDS-NA), industrial mechanisation and automation

Philips Research Laboratories, Eindhoven

The Philips Research Laboratories in Eindhoven, started in 1914, are the cornerstone of Philips Research. In 1989, the Laboratories occupy the buildings shown in the photograph. Important activities include exploratory research, computational physics, magnetic and optical recording, IC design and technology, television systems and computer architecture.



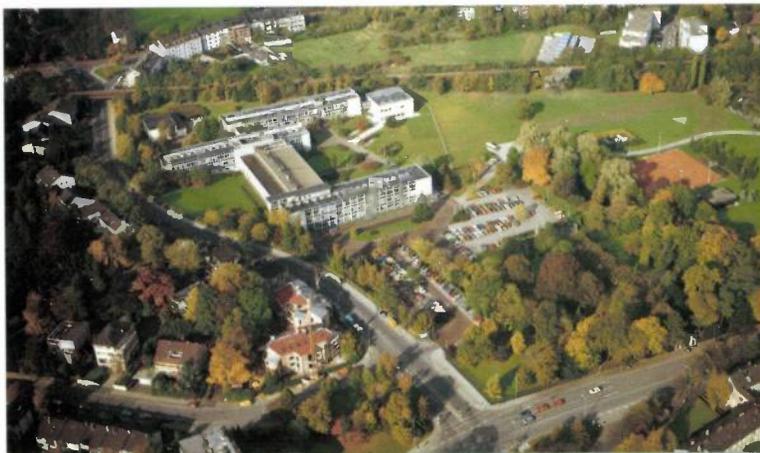
Eindhoven

Philips Research Laboratory Brussels, Brussels, Belgium

Software and information technology, image analysis, expert systems

Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany

Gas-discharge lamps, electric motors, CVD techniques for making glass optical fibres for communications, ceramic materials



Aachen



Brussels

**Laboratoires d'Electronique et de
Physique Appliquée,
Limeil-Brévannes, France**

IC technology and design in III-V
compounds, optoelectronics, digital
signal processing (for televi-
sion and infrared)



Limeil-Brévannes



Hamburg

**Philips GmbH Forschungslaboratorium Hamburg,
Hamburg, West Germany**

Communications and office automation, speech
recognition, imaging systems for medical diagnostics

**Philips Research Laboratories, Sunnyvale,
Sunnyvale, CA, U.S.A.**

The work here is concentrated on IC technology and design
in Si



Sunnyvale



Redhill

**Philips Research Laboratories,
Redhill, Surrey, England**

Solid-state electronics for discrete
components in Si and GaAs, radar
and microwave systems, liquid-crystal
displays for television

Predicting the properties of materials: dream or reality?

M. F. H. Schuurmans, R. Coehoorn, R. Eppenga and P. J. Kelly

It has become standard practice in IC technology to simulate the operation of an IC before it is fabricated in silicon. The advantages are obvious: it saves money and time. Today, new materials — combinations of known semiconductors or magnetic materials — are often required for special applications. Here again time and money can be saved by calculating the properties of these new materials beforehand. Current developments in this relatively new field are discussed in the article below.

Why?

As requirements for materials become more varied and tests and experimentation become more time-consuming and expensive, the prediction of the properties of materials will gain in importance.

We shall consider this proposition as applied to semiconductors and magnetic materials, which are widely used in electronic components for radio, television, optical communications, and telephone and computer systems. Examples of semiconductor devices are transistors, light-emitting diodes and lasers. Magnetic materials are used in loudspeakers, motors and in magnetic recording.

A wide range of materials can be made today by using special growth techniques. Monolayers of atoms of any kind can be deposited on a substrate. Almost any chemical composition can be made in this way. Examples of such growth techniques are Molecular Beam Epitaxy (MBE) and Metal Organic Chemical Vapour Deposition (MOCVD). In MBE atoms are deposited on the substrate by evaporation from atomic sources in ultra-high vacuum. In MOCVD atoms are deposited on the substrate from a reactive gas mixture. Both techniques can be used to make a semi-

conductor laser that emits light at a single desired wavelength. One important application found in the Compact Disc player is the semiconductor laser, a layered structure containing Al, Ga and As. However, the correspondence between chemical composition and wavelength is not clear at the outset. The usual approach was purely empirical, a matter of trying and seeing whether it worked. But an MBE equipment might cost more than a million dollars, and because of contamination problems it can only be used for one semiconductor material, i.e. a single chemical compound. While MOCVD is not so expensive and is more versatile, investigating all the possible combinations of materials is still a hopeless task. Predicting the properties of semiconductors from their chemical composition and crystal structure is therefore very important.

Similar arguments apply to the prediction of some of the properties of metals and magnetic materials. Layered magnetic structures of almost any desired chemical composition can be made with the same special growth techniques, but they are expensive and take a long time.

Sometimes new materials with properties that can be determined from the known properties of the constituent materials will be of interest^[1]. For the

Prof. Dr Ir M. F. H. Schuurmans, Dr R. Coehoorn, Dr R. Eppenga and Dr P. Kelly are with Philips Research Laboratories, Eindhoven.

solid-state physicist working on the prediction of material properties this is the simplest case. However, the material is not really new, since the properties of the constituent materials can still be recognized. For 'really' new materials another approach is required. The last ten years have seen the development of the branch of computational physics known as *ab initio* calculations. They are called '*ab initio*' calculations ('*ab initio*' means 'from the beginning') because the calculations start from the most elementary information available: the atomic number — the number of electrons in the atom involved — and the crystal structure — the positions of the constituent atoms in the material. If any experimental information about the crystal structure of a material is available, it can be used. If the crystal structure is not known, it can be found from the calculations (the crystal structure giving the lowest energy in the ground state is the correct one; this is sometimes called the 'energetically most favourable structure'). We shall meet some examples of these calculations later. The atomic number of any atom can be found in the Periodic Table of the Elements (fig. 1). The constituents of the semiconduc-

The subject of this article is the *ab initio* calculation of material properties. After evaluating the complexity of this problem we shall discuss the density functional theory [2]. Most *ab initio* calculations for solids are based on this theory. This section can be omitted if the reader is not particularly interested in the technical details. We shall then discuss the usefulness of the *ab initio* calculations in predicting the properties of semiconductors and magnetic materials. Finally, we shall draw some conclusions on the extent to which the properties of materials can now be predicted and the success of this approach in theoretical solid-state research.

How?

So which properties of materials be predicted and how? To answer this question we first have to see how complicated the problem is. A cube of 1 cm³ of a material contains about 10²³ atoms. Each atom consists of a positively charged nucleus and a number of negatively charged electrons. The first problem is that in the material the electrons and the nuclei are all moving. Since nuclei are so much heavier than electrons (about 10 000 times) we can generally assume (but by no means always) that the two motions do not affect one another. So let us assume that the electrons move and the nuclei are stationary. The interaction between the electrons is now the biggest problem. In fact this problem cannot be solved exactly, even for a single atom (unless it is a hydrogen atom). There are a number of useful approximations that can be made, however. If the positions of the nuclei form a regular pattern (the crystal structure), we can reduce the interaction between 10²³ electrons to the interaction between the electrons that belong to just a few atoms (about 100) — provided, that is, that we are dealing with perfect crystals. In alloys, where the positions of the atoms do not form a regular crystal lattice, the problem is much more complex, and we shall not consider it further.

We have to remember here that the interaction between electrons in a material cannot be described by the laws of classical mechanics. We have to appeal to the laws of quantum mechanics, introduced by Bohr, Einstein and Schrödinger in the early 1900s. The interactions between the electrons in a material are described by the Schrödinger equation. From now on we

The Periodic Table of the Elements

1																	2
H																	He
3	4															10	
Li	Be															Ne	
11	12															18	
Na	Mg															Ar	
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102		
Fr	Ra	Ac															
Lanthanide series		Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu		
Actinide series		Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No			

Fig. 1. All the input parameters for *ab initio* calculations are given in this table: the Periodic Table of the Elements. The most important semiconductors are formed from elements in the 'red' part of the Table; the most important metals (for use in magnets) are elements from the 'green' part of the Table or their alloys.

tors that are of most interest for applications are found in the part of the Periodic Table shown in red. The actual semiconductors of most interest (besides silicon, of course) are GaAs and AlAs, which are used in solid-state lasers and high-speed transistors. Metals of interest are often alloys of elements found in the 'green' part of the Periodic Table. Elements that are well known as magnetic materials are Fe, Co and Ni. The best permanent magnets are based on the compounds Nd₂Fe₁₄B, SmCo₅ and Sm₂Co₁₇.

[1] R. Eppenga and M. F. H. Schuurmans, Theory of the GaAs/AlGaAs quantum well, Philips Tech. Rev. 44, 137-149, 1988.
 [2] J. P. Dahl and J. Avery (eds), Local density approximations in quantum chemistry and solid state physics, Plenum, New York 1984;
 N. H. March and S. Lundquist (eds), The inhomogeneous electron gas, Plenum, New York 1984.

have to bear in mind that the solutions of the Schrödinger equation are states and the corresponding energies or energy levels. A material can be in one of several states, each corresponding to a different energy. At low temperatures and if it is unaffected by external electric or magnetic fields, a material will be in the state with the lowest energy, the ground state (*fig. 2*). By applying a voltage it is possible to cause

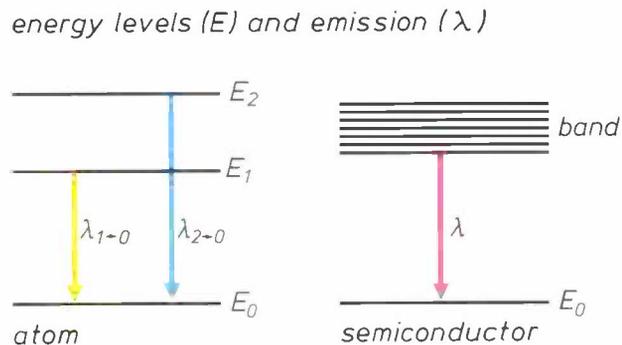


Fig. 2. The energy levels for a single atom and a hypothetical semiconductor material. The levels are indicated by horizontal lines and only give the magnitude of the energy. The set of energy levels is called the energy-level diagram. The ground state of a material is the one with the lowest energy, denoted by E_0 . The first excited state corresponds to the next higher energy, denoted by E_1 . The emission of light, at a wavelength λ , can take place when a material in an excited state returns to the ground state. This is represented symbolically by an arrow. In the semiconductor material energy levels may be grouped in bands. A diagram of this type is also called the band structure^[3].

the material to go into a state with a higher energy (the energy difference is supplied by the battery). This is an excited state. The emission of light (not necessarily visible) can take place when the material returns to the ground state. The wavelength of the light emitted is precisely determined by the difference between the energies in the excited state and the ground state (conservation of energy). The energy-level diagram, representing the set of energy levels for the ground state and the excited states, is essential to the understanding of the emission of light from a material. Since energy levels often occur in groups or bands, we often speak of band structure rather than energy levels^[3].

Even with the fast computers we have today it is not possible to find a solution to the Schrödinger equation that will describe the interactions between 100 electrons. The breakthrough in this field is not primarily due to the increased speed and memory capacity of computers. For the particular class of properties of materials associated with the ground state (crystal

structure, lattice parameters, binding energy), the breakthrough has come about because of a theory developed in the sixties: the density functional theory. The *ab initio* calculations based on this theory will be discussed briefly in the next section.

For other properties of materials associated with the excited states, such as the light-emitting properties of a semiconductor, there is as yet no really tractable theory. *Ab initio* calculations of the band structure are then useful, because they give a first indication of the location of the energy levels. A more satisfactory theory^[4] has been developed recently, and we shall return to this briefly later.

We can already see that the prediction of the properties of materials is no longer a dream. However, there is much room for improvement, and calculations performed without considering the physical aspects of the problem should be viewed with some reserve. We shall return to this later.

We shall now look at the density functional theory at an elementary level. Readers who are more interested in the answer to the question in the title of this article than in the details of the theoretical physics may prefer to omit this section.

The density functional theory

In the early sixties, P. Hohenberg, W. Kohn and L. J. Sham developed the density functional theory, on which most of the present calculations of the properties of materials are based. This theory gives information about a number of the properties of the ground state of an atom, a molecule or, more generally, a piece of material. In principle, the ground state and its corresponding energy can be found by solving the Schrödinger equation, which describes the quantum mechanics of the electrons. However, because of the electron-electron interaction the equation cannot be solved directly. Hohenberg and Kohn considered a function of the electron density^[5] that had a minimum corresponding to the energy in the ground state. The function assumes this minimum value for the electron density of the ground state. Some time later Kohn and Sham realized that this minimum formally corresponds to a solution of the Schrödinger equation for a single electron in an *effective potential field defined by the many-electron problem*. This was a huge step. In simple terms, it means that we started with interacting electrons, which all had to be considered at the same time, whereas now we can consider a single electron influenced by an effective force due to the nuclei and the other electrons. This force can be mathematically described in terms of an effective potential. *Fig. 3* shows a simplified picture of this de-

scription for a molecule consisting of two atoms. It will be clear that density functional theory will make matters even simpler when we have to deal with many interacting electrons. We have to remember that this is a *hypothetical* single-electron problem. This will be particularly important when we come to discuss the band structure.

There is a price to be paid for this simplification, of course: nobody has so far been able to calculate the exact effective potential, which is a function of the electron density. However, this potential can be approximated by using the known properties of a spatially homogeneous electron gas, and in many cases the approximation is satisfactory. The approximated potential is still a function of the electron density.

The Schrödinger equation is solved in stages; first the electron density is estimated, then the Schrödinger equation is solved using the approximation for the effective potential based on the estimated electron density. This results in a new value for the electron density. The process of solving the Schrödinger equation and calculating the electron density is repeated until the calculated electron density is the same as the estimated electron density. Now that computers are so much faster and have so much memory capacity this can be done for materials with a complex structure. The essential breakthrough, however, came from the newly acquired physical understanding of the problem.

We have to remember here that the density functional theory does not give the actual ground state. We only find the energy and electron density of a piece of material in the ground state. A theory that gives an approximation to the ground state and also the excited states is the Hartree-Fock theory^[2] much used by chemists. This theory has been applied very successfully to atoms and small molecules, but not to pieces of material comprising many atoms. The effective potential in the single-particle Schrödinger equation in this theory does not depend on the electron density, but on a suitably chosen set of functions each of which is a solution of such a single-particle Schrödinger equation. The solution of the single-particle Schrödinger equation in the Hartree-Fock theory is therefore more difficult than in the density functional theory^[6].

In the next section we shall discuss a number of applications of the density functional theory. Three of these applications (defects in semiconductors, magnetic materials and the emission of light in semiconductors) relate to research at Philips Research Laboratories, Eindhoven. The other two applications (semiconductors and semiconductor surfaces) relate to work done elsewhere.

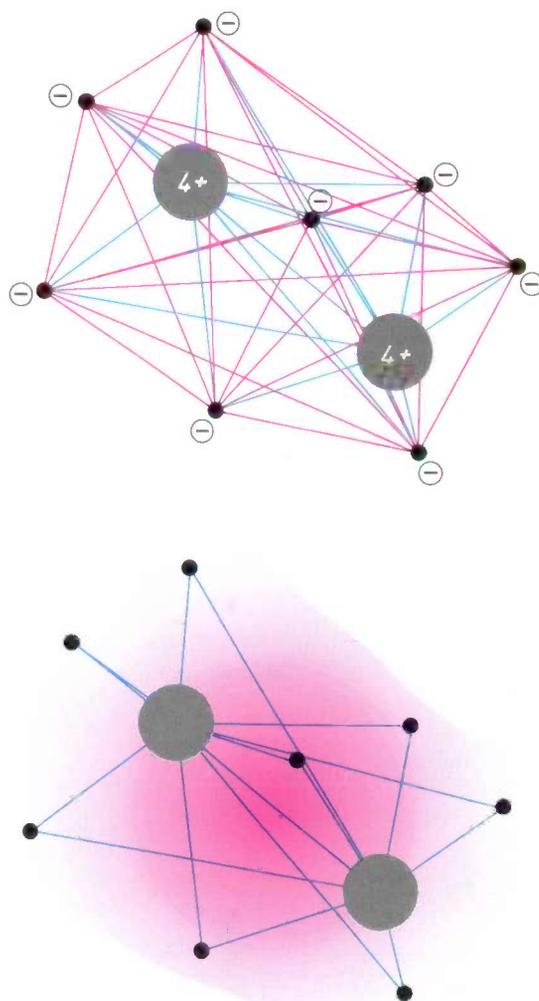


Fig. 3. An 'artist's impression' of the interaction between ions and electrons in a molecule consisting of two atoms. Interaction between each of the eight electrons and the two ions is indicated by blue lines. The mutual interaction between the electrons is indicated by red lines. If the interaction is described by the density functional theory the network of red lines disappears and the potential that a hypothetical electron is subjected to is indicated by the red background. The variation in the depth of colour corresponds to the variation of the potential.

[3] 'Band structures' are most frequently encountered in the context of single-electron energy-band diagrams. Fig. 2 shows a diagram for a large number of electrons.

[4] S. G. Louie, Quasiparticle energies in semiconductors and insulators, Proc. 18th Int. Conf. on the Physics of semiconductors, Vol. 2, Stockholm 1986, pp. 1095-1102; R. W. Godby, M. Schlüter and L. J. Sham, Accurate exchange-correlation potentials for semiconductors and their connections with the quasiparticle energies, Proc. 18th Int. Conf. on the Physics of semiconductors, Vol. 2, Stockholm 1986, pp. 1103-1106.

[5] Electrons move between the nuclei in a material. The electron density corresponds to the probability of finding an electron at a particular position in the material.

[6] A version of the Hartree-Fock theory related to the density functional theory is the Hartree-Fock-Slater theory. In Slater's version of the method, the effective potential from the Hartree-Fock theory is approximated by a function of the electron density. Unfortunately, this approximation is often an oversimplification and a correction parameter has to be used.

Prediction of properties of a material in the ground state

Semiconductors

What can we predict with the density functional theory? Let us first look at silicon, the most important material in the semiconductor industry. Silicon has various possible crystal structures, depending on the applied pressure. At room temperature and atmospheric pressure it has the crystal structure of diamond (see fig. 4). Does the density functional theory confirm this? It does indeed. We shall now discuss results mainly obtained by M. L. Cohen and his collaborators at the University of California, Berkeley [7].

The energy of silicon in the ground state can be calculated for different crystal structures and for different distances between the atoms in the lattice. It turns out that the diamond structure gives the lowest energy when the distance between the atoms in the lattice is within 1% of the measured lattice spacing. In addition, the compressibility of the material, that is the change in energy when the material is compressed, can be predicted to within 2% of the measured value.

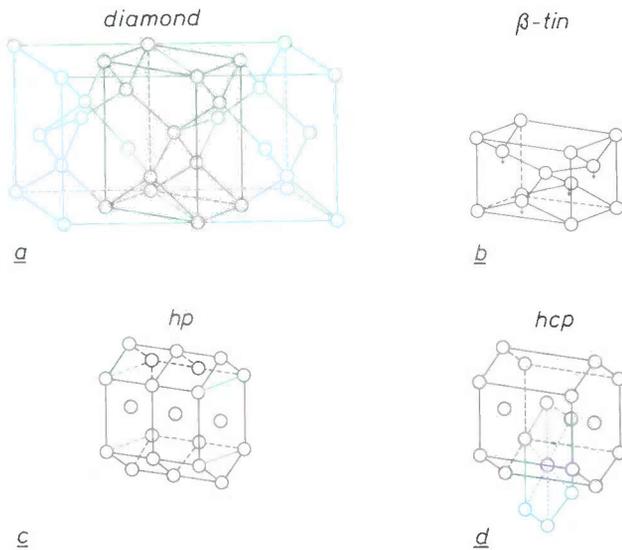
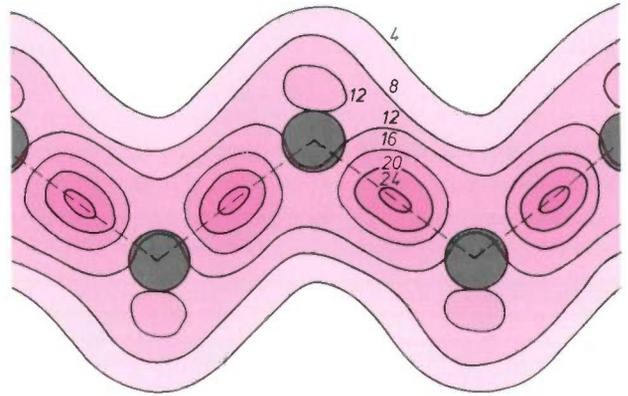


Fig. 4. Crystal structures of silicon, for increasing pressure. a) diamond structure b) β -tin structure c) primitive hexagonal crystal structure and d) hexagonal close-packed crystal structure.

The calculation also shows that above a critical value of the applied isotropic pressure (10 GPa) the diamond structure no longer has the lowest (free) energy [8] (see fig. 4a). The arrangement of the atoms with the lowest ground-state energy under these conditions can be seen in fig. 4b. The new crystal structure is that of β -tin. At still higher pressures we find transitions to the primitive hexagonal structure (fig. 4c) at 16 GPa and finally to the hexagonal close-packed

theory



experiment

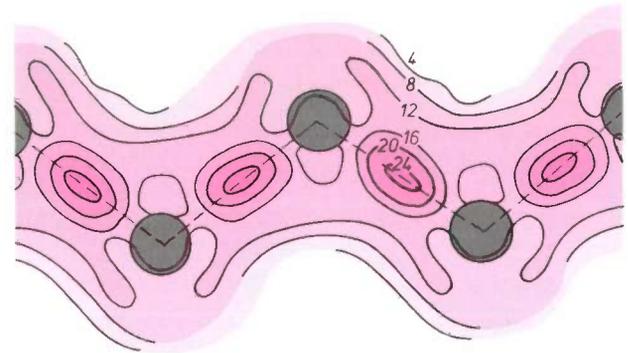


Fig. 5. Calculated and measured electron density in a (110) plane in the silicon crystal. Nuclei are indicated by grey circles and lines of constant electron density are shown. The high electron density between the nuclei corresponds to the bonds between the atoms in the crystal.

structure (fig. 4d) at 40 GPa. The predictions indicate that these modifications of silicon under pressure are metals. When silicon is subjected to increasing pressure, experiments do indeed reveal the predicted sequence of crystal structures. Transitions between the different structures occur at the predicted pressure values and under sufficient pressure silicon is indeed a metal. If the temperature is low enough it even becomes a superconducting metal. The calculated electron density of silicon in the ground state in the diamond structure is in excellent agreement with the measured density (see fig. 5). Note that these calculations do not involve free parameters. *The calculations really are 'ab initio'*. The only input is the atomic number of silicon from the Periodic Table: 14.

Semiconductor surfaces

A good understanding of semiconductor surfaces and interfaces between semiconductors is very important for the continuing development of IC technology. *Ab initio* calculations based on the density functional

theory can contribute substantially here. We shall discuss one spectacular example: the (111) surface of silicon.

The atoms at the surface of a semiconductor and those just beneath it are not generally in the positions that might be expected from the periodic arrangement of the atoms in the semiconductor. This behaviour can be explained in the following way. Atoms in the bulk of a semiconductor are bound to each other because the four electrons most weakly bound to the atoms are distributed over the four orbitals pointing to the nearest-neighbour atoms. An atom at the surface, however, has fewer than four neighbours and thus has unbonded or 'dangling' orbitals. These dangling orbitals do not point at neighbouring atoms but into the empty space beyond the semiconductor surface. Reordering of the atoms at the surface and those just beneath it, with a partial reorientation of the orbitals, may reduce the distance between the dangling orbitals. The corresponding surface reconstruction, i.e. the new periodic arrangement of the atoms at the surface, results in a lower energy for the surface. The surface reconstruction with the lowest ground-state energy can be found with the aid of density functional theory.

Low-energy electron diffraction (LEED) at the (111) surface of silicon formed by cleaving a crystal shows that the surface has been subjected to a '(2x1)' reconstruction. The notation (2x1) means that the period of the atomic arrangement at the surface in one direction has become twice as large as the period in the bulk of the semiconductor and that it has remained the same in another direction. In the early eighties, Haneman's 'buckling' model^[9] was the accepted model for the silicon (111) surface; see fig. 6a. In this model, rows of atoms are positioned alternately higher and lower than they would be at a non-reconstructed surface. Dangling orbitals do not link up to the orbitals of the nearest-neighbour atoms, but with those of the next-nearest atoms. Some doubt about the correctness of this (2x1) reconstruction of the silicon (111) surface was expressed in 1982, when K. C. Pandey^[10] showed, with the aid of density functional theory calculations of the ground-state energies, that a model he had invented, the π -bonded-chain model, had the lowest ground-state energy, rather than the buckling model. In Pandey's (2x1) reconstructed model the topology of the atoms has been completely changed; see fig. 6b. The dangling bonds do now link up to those of the nearest neighbours. Support for Pandey's model soon appeared from various experiments^[11]. If the silicon is heated, the (2x1) reconstructed surface changes into a (7x7) reconstructed surface, now the subject of intense study with the

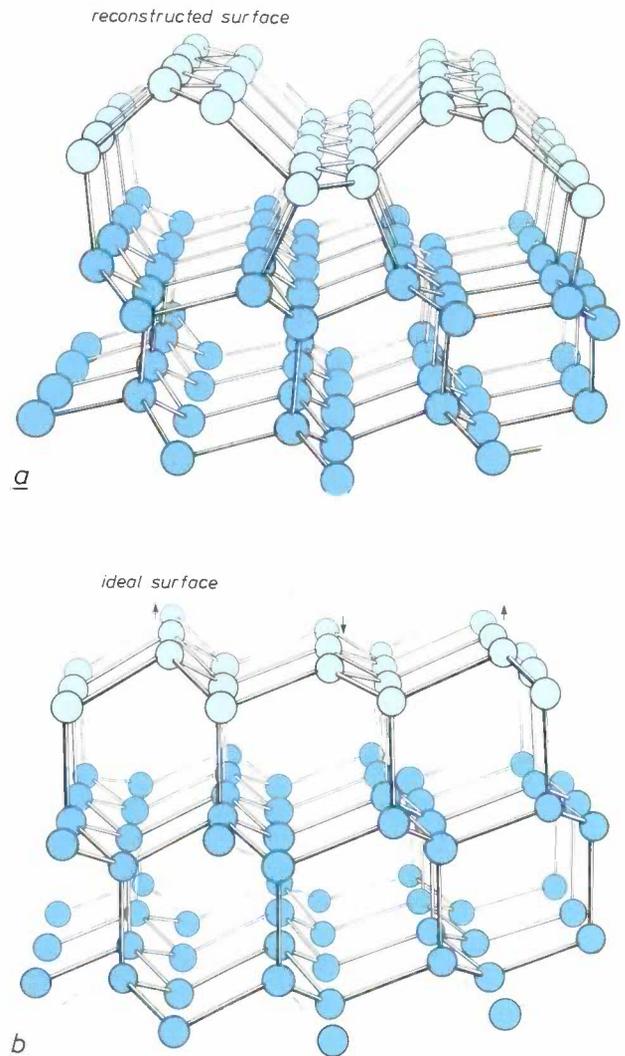


Fig. 6. a) The buckling model of the (111) surface of silicon. As a result of the reorientation of the surface atoms rows of atoms are positioned alternately higher and lower than in the structure in the bulk of the material. The surface atoms are shown in a lighter blue than the atoms in the bulk. b) π -bonded chain model of the (111) surface of silicon. As a result of the displacement of the surface atoms, alternating pentagonal and heptagonal rings occur. The normal hexagonal structure can be seen in the bulk of the material.

- [7] M. L. Cohen, Pseudopotentials and crystal structure, in: Structure and bonding in crystals, Vol. 1, M. O'Keeffe and A. Navrotsky (eds), Academic Press, New York 1981, pp. 25-48.
- [8] D. Erskine, P. Y. Yu, K. J. Chang and M. L. Cohen, Stability of the Si lattice: electron-phonon interactions and superconductivity in hexagonal Si, Proc. 18th Int. Conf. on the Physics of semiconductors, Vol. 2, Stockholm 1986, pp. 1217-1220.
- [9] D. Haneman, Surface structures and properties of diamond-structure semiconductors, Phys. Rev. **121**, 1093-1100, 1961.
- [10] K. C. Pandey, Reconstruction of semiconductor surfaces: buckling, ionicity, and π -bonded chains, Phys. Rev. Lett. **49**, 223-226, 1982.
- [11] R. M. Tromp, L. Smit and J. F. van der Veen, Si(111)-(2x1) surface: buckling, chains or molecules?, Phys. Rev. Lett. **51**, 1672-1675, 1983; P. Chiaradia, A. Cricenti, S. Selci and G. Chiarotti, Differential reflectivity of Si(111) 2x1 surface with polarized light: a test for surface structure, Phys. Rev. Lett. **52**, 1145-1147, 1984; R. M. Feenstra, W. A. Thompson and A. P. Fein, Real-space observation of π -bonded chains and surface disorder on Si(111) 2x1, Phys. Rev. Lett. **56**, 608-611, 1986.

scanning tunnelling microscope^[12]. The (7x7) reconstruction is unfortunately too complicated for *ab initio* calculations on present-day computers. In a few years, however, the situation may have changed.

Defects in crystal lattices

In the third example of *ab initio* calculations of properties of the ground state we leave materials that have a perfect periodicity (in two directions at the surface, in three directions in the bulk of the material) and we discuss as an example of an impurity (a defect) in a semiconductor some aspects of oxygen in silicon.

The importance of oxygen-related defects in silicon becomes clear when we remember that silicon dioxide, a stable insulating material, is often used in IC technology. As a result of various heat treatments in the manufacturing process, oxygen will always be present in substantial concentrations in silicon, and either on its own or in combination with other impurities gives rise to a large variety of defects. This is also true for nitrogen and carbon. Because these defects may affect the conducting properties of silicon, it is important to understand the physical properties of these defects.

In general, it is not possible to determine experimentally either the position in the lattice of impurity atoms or the positions of the surrounding silicon atoms. Reliable calculations are necessary to complement or confirm the available experimental information. The calculations^[13] that we shall now look at briefly were performed at Philips Research Laboratories, and use the most advanced methods available today. They are made in the following way. First of all, the properties of the perfect silicon lattice are calculated. The foreign atom, the oxygen atom, together with the silicon atoms that are displaced as a result of imbedding the oxygen atom, is considered to be a spatially limited disturbance. Since the oxygen atom does not fit into the silicon lattice very well, about ten silicon atoms are displaced significantly. It is because this number is relatively large that it is difficult to find a solution to the problem.

One of the impurity configurations we investigated is an oxygen atom at an interstitial site in the diamond lattice of silicon. *Fig. 7a* shows one of the possible stable configurations in the (110) plane. We would expect the two neighbouring silicon atoms to be pushed aside to make room for the oxygen atom. This does in fact follow from the calculation of the ground-state energy as a function of the Si-O-Si bonding angle and the Si-O distance, but only if the displacements of the silicon atoms further away from the defect are included in the calculations. It turns out that in the configuration with the lowest energy the bond angle and distance are very similar to those found in the in-

termediate configuration, *fig. 7b*, the energy barrier to be overcome is the difference in energy between

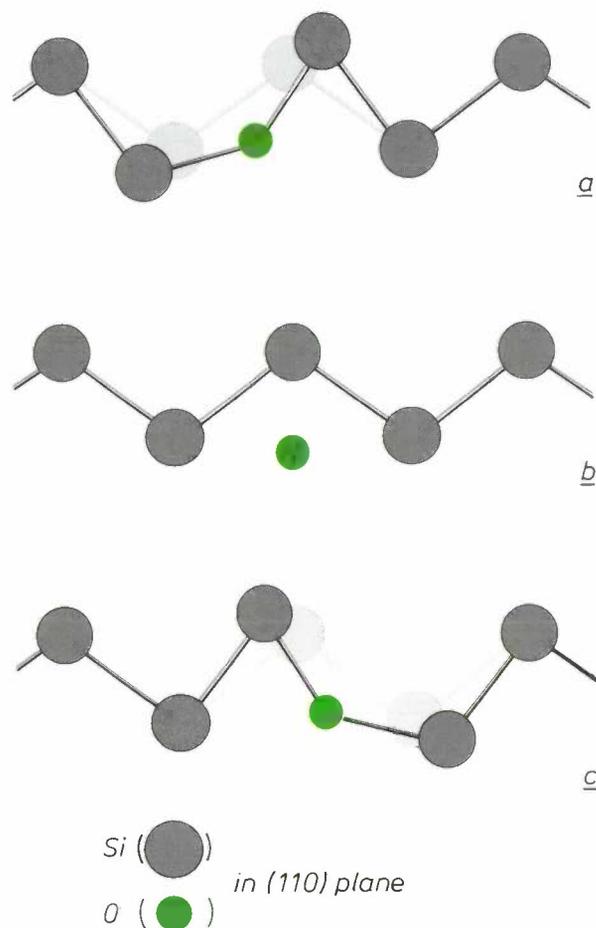


Fig. 7. Chain of silicon atoms (grey) in the (110) plane with an oxygen atom at an interstitial position. The oxygen atom does not fit into the silicon crystal lattice: the undisturbed positions of the silicon atoms are indicated by light-grey circles. Configurations *a* and *c* have the same energy, which is lower than that of configuration *b*. The calculated energy difference is 2.2 eV. This is an estimate for the activation energy of oxygen diffusion in silicon (the experimental value is 2.56 eV).

the configurations of *fig. 7b* and *7c*. The calculated energy barrier is 2.2 eV, in reasonable agreement with the experimental value of 2.56 eV. This is also the activation energy for the diffusion of an oxygen atom through a silicon crystal and it is therefore very important for the 'energy housekeeping' of oxygen in silicon. Calculations for defects consisting of several oxygen atoms are at present being carried out.

Magnetic materials

We now come to our last example in this section, magnetic materials. Every electron has an intrinsic angular momentum, the electron spin discovered by the Dutch physicists Goudsmit and Uhlenbeck, and the tiny circular current associated with it makes every electron a little magnet. The electron spin in a magnetic field can have either of two directions, which we shall call spin up and spin down. In non-magnetic materials there are equal numbers of spin-up and spin-down electrons. The magnetic moments of the electrons cancel exactly. In ferro- or ferrimagnetic materials the number of electrons with spin up is not the same as the number with spin down. The magnetic moment is this difference multiplied by the magnetic moment per electron, μ_B , the Bohr magneton.

The density functional theory can be extended to describe spin-dependent electron densities so that *ab initio* calculations of the ground-state properties of magnetic materials can be made. For example, calculations have been made to find out whether the spins of a magnetic material in the ground state are arranged in a ferromagnetic or an antiferromagnetic order [14]. The results are in good agreement with the available experimental data. The magnitude of the magnetic moments can also be compared with experiment. In ferromagnetic materials the magnetic moments of all the atoms point in the same direction. Examples are iron, cobalt and nickel. The calculated values of $2.15 \mu_B$ for iron, $1.54 \mu_B$ for cobalt and $0.61 \mu_B$ for nickel are in reasonable agreement with the experimental results 2.23 , 1.72 and $0.61 \mu_B$, respectively. The agreement is even better when the magnetic moment of the electrons in their orbital around the nucleus of the atom (a relativistic effect of about $0.1 \mu_B$) is taken into account.

We have already mentioned the permanent-magnet materials $\text{Nd}_2\text{Fe}_{14}\text{B}$, SmCo_5 and $\text{Sm}_2\text{Co}_{17}$ because of their important applications in loudspeakers and motors. The quality of a permanent magnet can be expressed in terms of a number of parameters (BH_{max} -product, remanent magnetization, coercive force) [15] and depends on the microstructure. Magnetic materials consist of small domains with the magnetizations pointing in different directions. The highest attainable value of the $(BH)_{\text{max}}$ -product is found to be proportional to the square of the magnetic moment in perfectly crystalline material and can therefore be calculated. The remanent magnetization and the coercive force are more difficult to calculate, because they depend on the domain walls and their displacements under the influence of a magnetic field.

In the search for new materials for permanent magnets attempts are often made to replace some or all of

the cobalt atoms in lanthanide-cobalt compounds by iron atoms, which have a slightly larger magnetic moment per atom. *Ab initio* calculations of the magnetic moments of the hypothetical compounds can give some guidance here. We shall now briefly describe some calculations carried out at Philips Research Laboratories [16]. Fig. 8 shows the magnetic moment of several compounds of yttrium and iron as a function of the percentage of iron. Yttrium is chemically equivalent to the lanthanides Sm and Nd, but has no magnetic moment; calculations therefore show the contribution of the iron atoms to the magnetic moment. The figure shows experimental and calculated [16] data. The contribution mentioned above from the orbital magnetic moment ($0.1 \mu_B$) is not in-

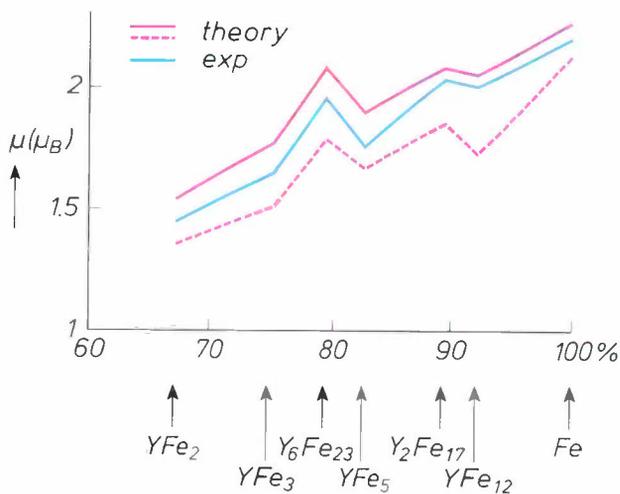


Fig. 8. Measured and calculated values of the average magnetic spin moments μ per iron atom for a series of compounds of iron and yttrium. The measured values are indicated by the blue line; the red lines refer to the calculated values (continuous red line for experimentally determined crystal volume, dashed red line for calculated crystal volume).

[12] I. K. Robinson, The role of strain in $\text{Si}(111) 7 \times 7$ and related reconstructed surfaces, *J. Vac. Sci. & Technol. A* **6**, 1966-1970, 1988.

[13] P. J. Kelly, Equilibrium geometries and electronic structure of oxygen related defects in silicon, *Proc. 15th Int. Conf. on Defects in semiconductors*, Budapest, 1988.

[14] J. Kübler, A. R. Williams and C. B. Sommers, Formation and coupling of magnetic moments in Heusler alloys, *Phys. Rev. B* **28**, 1745-1755, 1983.

[15] The $(BH)_{\text{max}}$ -product is proportional to the amount of magnetic energy that can be stored in the magnetic field around a permanent magnet. The remanent magnetization is the magnetization that remains in the material after switching off a strong magnetic field. The coercive field is the magnetic field required to demagnetize the material.

[16] R. Coehoorn, Electronic structure and magnetization in stable and hypothetical Y-Fe compounds, *Proc. ICM, J. Phys.*, Paris 1989.

cluded in the calculations. It should also be noted that the experimental data for the compounds YFe_5 and YFe_{12} were obtained by interpolation and extrapolation, respectively; these compounds are not stable in the crystal structure assumed. The trend of the experimental results is correctly predicted. The measured values for the magnetic moments lie between the values calculated with the experimental crystal volume and the calculated crystal volume (the difference between the experimental and theoretical curves is generally less than 10%).

In concluding this section we note that density functional theory can also be applied to atoms and molecules. It is found that the bonds between the atoms in a molecule can be described more accurately or more easily with this theory than with the existing methods most commonly used by chemists, namely the Hartree-Fock theory. A well-known example is the bond in the molybdenum dimer, Mo_2 . The Hartree-Fock theory verified the measured bond lengths, but only after very extensive computer calculations^[17]. Density functional theory required far less calculation for the same result^[18].

Prediction of the light emission of materials

When considering the application of materials in LEDs, lasers and television tubes, the electronic structure is extremely important. Can we predict it? Not yet, it seems: density functional theory only gives information about properties of materials in the ground state. This theory makes use of the Schrödinger equation for a single-electron problem, and a band structure can be calculated from this equation. However, this band structure has no real physical significance. In the density functional theory a hypothetical single-electron problem is solved to determine properties of the material in the ground state. States and energies derived for this hypothetical electron in the material are not necessarily related to the states and energies of a real electron. Predictions of the emission wavelength of silicon from this theoretical structure are therefore not correct; the calculated wavelength exceeds the experimental value by 40%. For metals, however, there is often good agreement between the calculated and experimental band structures. Some of the bands in the energy-level diagram for a semiconductor are in fact described very well by the theory. Because of this, some calculations of band structures have been made with the density functional theory, even though there was no proper theoretical or physical basis for using it in this case. Such calculations are called *ab initio* band-structure calculations. Well-known problems^[19] such as the monoxides of nickel

and manganese show that caution is indicated, however. These materials are good electrical insulators but the theory indicates that they ought to be conductors. Otherwise the properties of these monoxides in the ground state are predicted correctly.

Another aspect is that of the accuracy of the calculations. Comparison of experimental and theoretical band structures and assessment of likely applications for these materials in lasers and LEDs demands an accuracy that the density functional theory cannot provide. In such cases semi-empirical theories, where experimental information is used to meet the desired accuracy, may be of use.

Some calculations at Philips Research Laboratories^{[20][21]} of the light emission from superlattices consisting of a periodic structure of alternating thin layers of AlAs and GaAs (see *fig. 9*) may serve as an example. Nowadays very thin layers of these semiconductor materials can be successfully grown with a perfect lattice match by using MBE and MOCVD. The properties of superlattices can be very different from the properties of the semiconductor materials from which the superlattice is constructed. The semiconductor GaAs has a strong infrared emission ($\lambda = 870$ nm). On the other hand the semiconductor AlAs has a weak emission in the yellowish-green part of the visible spectrum ($\lambda = 570$ nm). How strong an emission will GaAs/AlAs superlattices give, and at what wavelength? This will depend on the thickness of the GaAs and AlAs layers, of course. If

- [17] M. M. Goodgame and W. A. Goddard, III, Nature of Mo-Mo and Cr-Cr multiple bonds: a challenge for the local-density approximation, *Phys. Rev. Lett.* **48**, 135-138, 1982.
- [18] B. Delley, A. J. Freeman and D. E. Ellis, Metal-metal bonding in Cr-Cr and Mo-Mo dimers: another success of local-spin-density theory, *Phys. Rev. Lett.* **50**, 488-491, 1983; J. Bernholc and N. A. W. Holzwarth, Local spin-density description of multiple metal-metal bonding: Mo_2 and Cr_2 , *Phys. Rev. Lett.* **50**, 1451-1454, 1983.
- [19] J. H. de Boer and E. J. W. Verwey, Semi-conductors with partially and with completely filled $3d$ -lattice bands, *Proc. Phys. Soc. (London) A* **49**, extra part (31 August 1937), 59-71, 1937; J. Zaanen, G. A. Sawatzky and J. W. Allen, Band gaps and electronic structure of transition-metal compounds, *Phys. Rev. Lett.* **55**, 418-421, 1985. The high- T_c superconductors are a recent example where problems are encountered if the density functional band structure is used naively, see for example M. F. H. Schuurmans and L. F. Feiner, Naar een theorie van de nieuwe supergeleiders, *Ned. T. Natuurk. A* **54**, 71-76, 1988.
- [20] R. Eppenga, M. F. H. Schuurmans and S. Colak, New k,p theory for GaAs/Ga_{1-x}Al_xAs-type quantum wells, *Phys. Rev. B* **36**, 1554-1564, 1987; For more experimental data see: K. J. Moore, G. Duggan, P. Dawson and C. J. Foxon, Short-period GaAs-AlAs superlattices: optical properties and electronic structure, *Phys. Rev. B* **38**, 5535-5542, 1988.
- [21] R. Eppenga and M. F. H. Schuurmans, Thin [001] and [110] GaAs/AlAs superlattices: distinction between direct and indirect semiconductors, *Phys. Rev. B* **38**, 3541-3544, 1988.
- [22] R. W. Godby, M. Schlüter and L. J. Sham, Quasiparticle energies in GaAs and AlAs, *Phys. Rev. B* **35**, 4170-4171, 1987.
- [23] P. Soven, Coherent-potential model of substitutional disordered alloys, *Phys. Rev.* **156**, 809-813, 1967.

both layers are relatively thick (> 40 nm), properties of both GaAs and AlAs will be observed. If the layers are thinner (2-40 nm), the constituent materials will affect one another. Experimentally, a strong emission is found, which originates from the GaAs, but at a wavelength shorter than that normally associated with GaAs and longer than that normally associated with

light emission as well, is now gaining ground^[4]. The equations in this theory are very similar to the equations in the density functional theory. This means that density functional theory can be of real benefit in making exact calculations of band structure. So far the new theory has only been used for simple problems^[22], but further developments are likely.

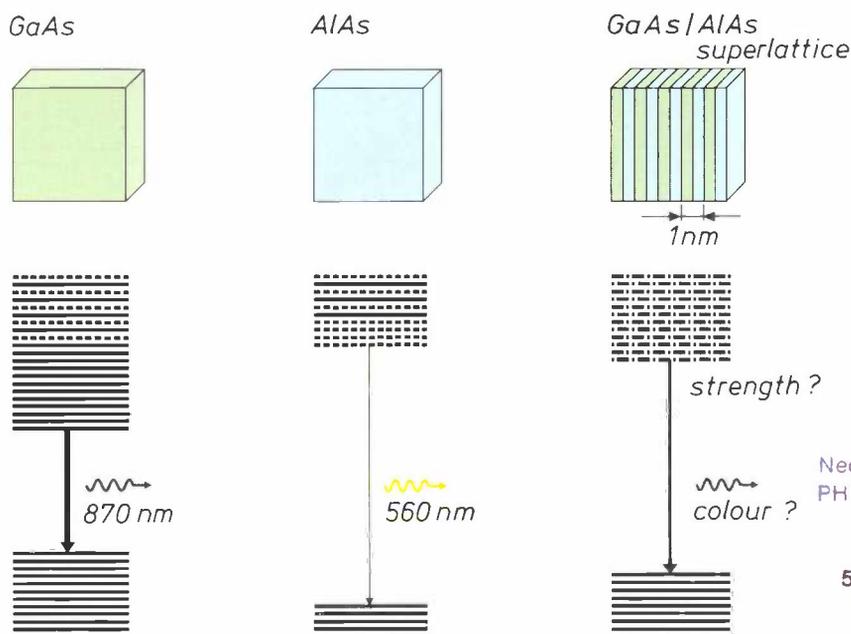


Fig. 9. GaAs has a strong emission in the infrared (indicated by a thick arrow and a black 'wave'). AlAs has a weak emission in the yellowish-green part of the spectrum (indicated by a thin arrow and a yellowish-green wave). *Ab initio* calculations help to provide values for the intensity and wavelength of the emission from a superlattice consisting of alternate layers of GaAs and AlAs.

AlAs. Semi-empirical calculations^[20] based on the properties of AlAs and GaAs for a superlattice of GaAs and AlAs layers of thickness 4 nm indicate that there should be a strong emission of red light. This is in agreement with the experimental results.

If the layers are even thinner (< 1 nm), semi-empirical calculations are no longer permissible, since we now have a truly new material. *Ab initio* band-structure calculations^[21] show that weak emission would be expected from a superlattice consisting of GaAs and AlAs layers of thickness between 0.3 and 1.1 nm. When corrections based on experimental data for AlAs, GaAs and AlGaAs have been made, the calculations show that orange light will be emitted. A comparison with experimental data from perfect superlattices confirms our results.

A theory that gives a correct description of experimentally observed band structures, and therefore the

Dream or reality?

Can we predict the properties of materials correctly yet? The preceding sections show that any answer here must be qualified. For example: predictions of properties of materials in the ground state are realistic, properties associated with excited states should be considered with some reservation. Moreover, the theory can only be applied to perfect crystals. Alloys present us with new problems, though there are methods for dealing with them^[23]. When we have invented a new material, we often do not know the positions of the constituent atoms very accurately. We can only make further progress if the material has a definite symmetry.

So is the enthusiasm with which *ab initio* calculations have been welcomed in recent years justified? Yes, it is. We can now see the way ahead to making real predictions. Trial calculations of structures con-

Ned. Philips Bedrijven B.V.
PHILIPS RESEARCH LABS.
LIBRARY WY - 1
P.O. Box 80.000
5600 JA EINDHOVEN
THE NETHERLANDS

sisting of very thin layers of semiconductor or magnetic materials, or combinations of these, which may be of interest to industry now or in future, have already proved useful^[24]. The same is true for calculations of bond lengths and angles in large molecules now under way in the petrochemical and pharmaceutical industries. In the preceding sections we have shown that these should be viewed with some reservation. We are still a long way from a situation where service centres can conjure up information about a material from a computer. What we really need now is good team work between 'computational' specialists, physicists and chemists. This is the only way to obtain reliable results. The divide between computational physics and phenomenological physics that some think will be reflected in *ab initio* calculations is therefore undesirable, at least for the time being. And what is more, the question as to whether calculations should be based on the Hartree-Fock theory or on the density functional theory is misplaced. The information from the two approaches is complementary — certainly for atoms and molecules: energy schemes and states from the Hartree-Fock theory and ground-state properties from both theories.

^[24] Such calculations are now being done in the Advanced Theoretical and Experimental Physics group at Philips Research Laboratories in Eindhoven. Calculations of the magnetic anisotropy of magnetic multilayers are currently being explored.

^[25] See for example: A. R. Miedema, The atom as a metallurgical building block, Philips Tech. Rev. 8, 257-268, 1978/79.

It should perhaps be noted here that larger computers *alone* will not give the answers to the questions we are asking. Much will depend on a better understanding of the physics of interactions between larger numbers of particles. But this in no way removes the necessity for faster computers of greater capacity for future developments in the prediction of properties of materials.

Finally, it should be remarked that the *ab initio* calculation of the properties of materials will be very demanding. It is therefore still useful to develop models^[25] for extrapolating properties of materials from experimental data for other materials, since such models can reduce the need for calculation and experiment.

Summary. Before new materials are used in commercial products their properties must be known. This information can be acquired from experiments, of course. Properties can also be calculated even before the materials exist. One of the theories that can be used in these calculations is the density functional theory. The present developments in this field are discussed with the aid of results obtained from the theory. These results relate to materials in the ground state: the different crystal structures of silicon at increasing pressure, the rearrangement of atoms at the surface of silicon, the consequences for the crystal lattice when oxygen is incorporated and the magnetic behaviour of iron-yttrium compounds. The theory can also be used, with some modification, for predicting the emission of light from GaAs/AlAs superlattices.

A sound basis for the generation of explanations in Expert Systems

A. Bruffaerts, E. Henin and A. Pirotte

From the very first days of the computer, non-numerical applications such as theorem proving and problem solving have been subjects of research. It is only during the last decade that a wider commercial interest has developed in artificial intelligence, a sector of informatics in which computer programs known as expert systems have been developed for problem solving that requires human expert behaviour. If expert systems are to be trusted by their users, then they should not only be able to provide a solution to a problem, they should also be able to explain it and show that it is correct and adequate.

Introduction

Expert systems solve problems that require human expertise in specialized domains. They work by exploiting knowledge ordinarily acquired from human experts and simulate some of the functions of a human expert in the application domain. Expert-system techniques currently work well only in carefully delimited domains, with strict limitations on context, relevance, and 'common sense'. While the most enthusiastic advocates of the technology argue that this is only a provisional limitation^[1], others are more sceptical and warn of connotations of the term *expert* not warranted by the achievements or the prospects of the technology^[2].

Knowledge in expert systems resides in a *knowledge base (KB)* (from now on) in the form of more or less independent items of knowledge frequently called *rules*. In some systems, rules have a rather declarative interpretation close to the implication of logic (*IF premise THEN conclusion*), while, in others, the intended interpretation is more procedural (*IF condition THEN action*). Knowledge for solving a problem is invoked by a computer program called an *inference engine*. By examining data derived from a query and representing a problem being solved, the inference

engine selects appropriate rules in the *KB* and executes them on the data for the problem (see *fig. 1*).

The performance of an expert system depends on the adequacy of its knowledge and on how effectively this knowledge is used, hence the term *knowledge-based system*, often used as a synonym.

Expert-system technology is often used in applications where a substantial part of the knowledge is heuristic. The experts' understanding of the domain, although sufficient to address certain classes of problems adequately is changing and imperfect. Therefore, the knowledge represented in expert systems, like the experts' knowledge, varies in its depth, in its stability with time, and in the degree of confidence that experts have in it. Also, application domains of interest for expert systems are usually specialized technical domains where the ordinary user's grasp of the subject is far less detailed than the expertise of the specialist.

Because expert systems have these characteristics, simply stating the solution to a problem is usually insufficient to convince users of its adequacy. Further detail about the solution must be available, reasons should be given for it and it must be seen to be

Dr A. Bruffaerts, Dr E. Henin and Dr A. Pirotte are with Philips Research Laboratory, Brussels, Belgium. This work was partially supported by the Commission of the European Communities, under Project ESTEAM-316 of the ESPRIT Programme.

[1] E. A. Feigenbaum and P. McCorduck, *The fifth generation: artificial intelligence and Japan's computer challenge to the world*, Addison-Wesley, Reading 1983.

[2] T. Winograd and A. Flores, *Understanding computers and cognition*, Ablex Publishing, Norwood 1986.

justified. In human interactions, these activities are part of the work of the human expert and they correspond exactly to the ordinary notion of 'explanation': explaining is making something clear or giving reasons for it. If expert systems are to exhibit expert

Varieties of explanations

The most obvious users of explanations are the non-expert users of an expert system. For these users it is necessary to *justify the solutions*, i.e. convince them that the solutions are correct and adequate. This conviction will follow from: (1) a basic understanding of the domain of application, (2) the acceptance of the underlying hypotheses from which the system derives a solution, and (3) the realization that the steps in the reasoning of the derivation are reasonable.

The amount of justification depends on the application and on the users. A conventional computer program (i.e. not an expert system) that inverts a matrix, for example, will be trusted if its users are confident that it codes a classical, well-known and well-analysed algorithm for doing the work. On the other hand, the adequacy of the solution provided by a program which advises users on how to invest their savings cannot be characterized by universally accepted criteria or hypotheses. This difficulty is not due to the program itself nor to the fact that the task is accomplished by a program. It is inherent in a domain where few established hypotheses are accepted by all the human experts and where some of the steps in the experts' reasoning depends more on their experience in the domain than on causal principles. This kind of reasoning, referred to as heuristic reasoning, is an essential part of human expertise. 'Explanation' thus implies: (1) precisely stating the hypotheses behind the reasoning leading to a solution so that users may decide whether they agree with them and (2) showing that the derivation of the conclusions from the hypotheses is reasonable.

If the user is to understand a solution, then it is preferable for him to be able to discuss it and ask for alternative solutions. The non-expert user may also be interested in answers to general questions, not directly linked to a particular problem being worked on, about the application domain, about its terminology, its principles and the methods of reasoning. In this *educational role*, a system or an expert teaches the basics of the application domain to the user.

The system is intended for the end-users, of course, but the explanations produced by an expert system will also be of value to 'knowledge engineers' and experts in the application domain. Knowledge engineers interact with the expert system, first to create the initial version of the system, and later to control its operation and extend its capabilities. They have to monitor and debug the system at several levels. Explanations can provide confidence and make errors more apparent than just by tracing the execution of the program. Human experts are the source of the knowledge represented in the system. Explanations

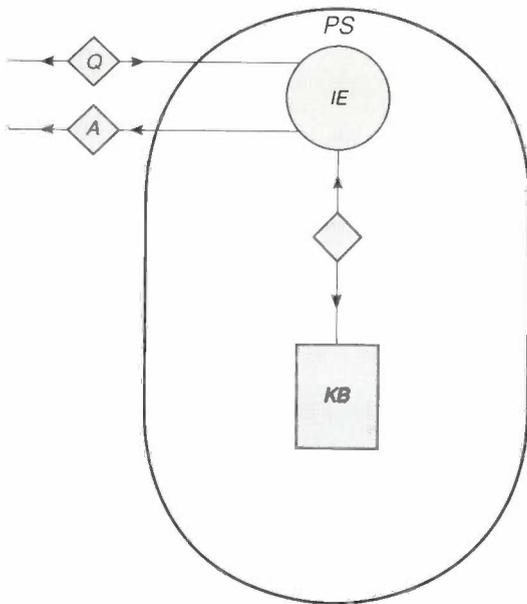
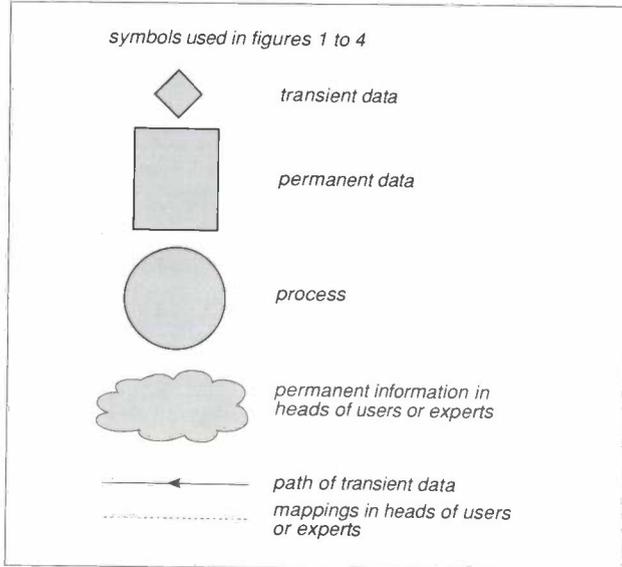


Fig. 1. Simple (pure) problem-solving by an expert system. The problem-solver *PS* in the expert system is addressed with a query *Q*. An inference engine *IE* selects and executes rules from the knowledge base *KB* to derive an answer *A*.

behaviour, they certainly ought to be able to explain how they arrive at their solutions and justify the methods they use to obtain them.

It the rest of this article we shall first discuss the kinds of explanations that are required, and who needs them and when, before we look more closely at the processes of generating them.

can give them a better mastery of the knowledge represented in the system and will help them to clear up misunderstandings. They can use the system to test the validity and compatibility of new items of knowledge or of strategies they are thinking of adding to the *KB*.

The main differences in the needs of the various kinds of users concern: (1) the level of detail at which explanations should be formulated and (2) the discrepancy between the model of the domain that the user has in mind and the model encoded by the expert system. These differences affect both the content and the structure of adequate explanations.

One difference concerns the degree of familiarity with the internal workings of the expert system. Most end-users will want to address the system and to receive its output in some kind of natural language, so as to minimize their interaction with the computational mechanisms of the system. The natural language can be, for example, a simple version of English adapted to the specific application domain. On the other hand, when knowledge engineers are monitoring the system they will want something more like an operational trace of its operation. They will want to know, say, which item of knowledge was activated at a given point in the formation of a solution.

Another difference concerns the level of understanding about the domain of expertise. Experts must be able to ask for concise explanations focused on particular items of knowledge, while a first-time user may need a detailed justification of the simplest answers, explicitly linked to the basic elements of the application domain. But even so, end-users would probably want the level of detail in the explanations to evolve as they become more expert themselves. Thus, a general requirement for a good explanation system is that it should deliver explanations that can adapt to the needs of different classes of users.

Research on the production of explanations addresses two questions: (1) what is the knowledge necessary for good explanations and (2) which tools and strategies will be adequate for representing that knowledge and for using it when generating explanations.

Early attempts at generating explanations

Explanations produced by early expert systems consist essentially of a trace of the steps leading to a solution^{[3][4]}, (fig. 2). These systems answer questions about the derivation of a solution by displaying the complete trace of rule applications in the process of solving the problem (*how-explanations*). Whenever they request additional information from the user, the systems can explain why the new information is nec-

essary and how it will be used, by displaying a trace of their current reasoning (*why-explanations*). These systems are also able to explain why they did not draw a conclusion that the user had in mind (*why-not-explanations*).

However, tracing the reasoning of the system merely describes its behaviour without fully justifying it. Besides accounting for a solution, adequate explanations should also be able to give the reasons for choosing the steps that led to it, by justifying these steps and their validity by references to general principles of the application domain.

Sometimes, the reasoning followed in solving a problem does not use strong causal arguments but relies instead on empirical associations or heuristics (rules of thumb) that the human expert has learned through experience. The *KB* then contains a representation of the heuristic methods that are more efficient because they short-circuit explicit causal reasoning. Such knowledge is sometimes called 'compiled' knowledge.

However, the reasoning reflected in a convincing explanation may need to be more detailed than that necessary for problem-solving alone. Some of the 'missing' steps must then be made explicit — provided

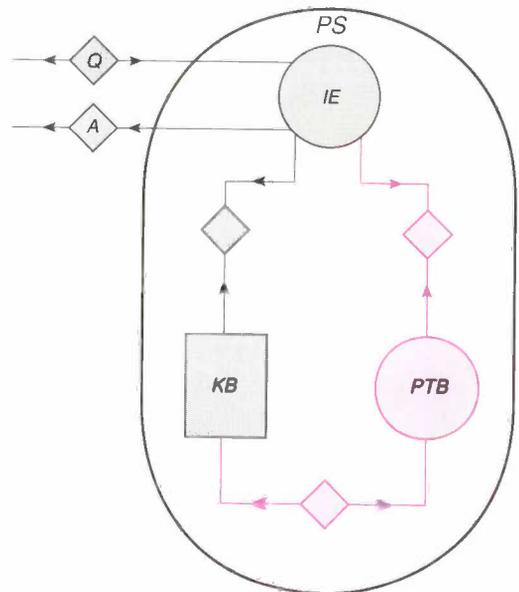


Fig. 2. Problem-solving with proof-tree generation. When a user of an expert system wants an explanation to go with an answer, a proof-tree builder *PTB* must be activated to trace the steps followed in deriving the answer. How-, why-, and why-not-explanations are produced from such a proof tree.

[3] B. G. Buchanan and E. H. Shortliffe, Rule-based expert systems: the MYCIN experiments of the Stanford heuristic programming project, Addison-Wesley, Reading 1984.

[4] P. Hammond, Micro-Prolog for expert systems, in: Micro-Prolog: programming in logic, K. L. Clark and F. G. McCabe (eds), Prentice-Hall International Series in Computer Science, Englewood Cliffs 1984, pp. 294-315.

the corresponding knowledge is available, of course. With the passage of time, researchers have come to realize that more is required than purely performance-oriented knowledge, and that some knowledge must also be represented in the *KB* for giving explanations.

This trend towards the incorporation of extra explanatory knowledge in the *KB* runs in parallel with the evolution of expert-system technology in such a way that there is a clear distinction between different

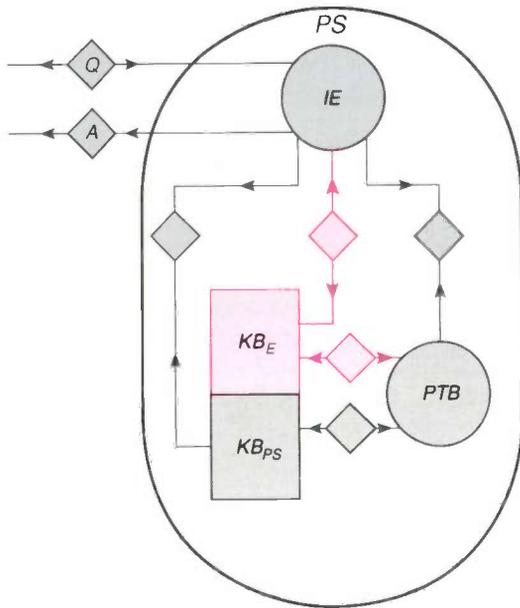


Fig. 3. Problem-solving with solution justification. Simply displaying the trace of rule applications is not satisfactory for all users in all circumstances. Accounting for and justifying the steps in the process of solving the problem may require extra knowledge KB_E in the knowledge base.

types of knowledge to make the *KB* easier to use and modify. Various types of knowledge come into the picture:

- strategic knowledge or knowledge about domain principles which explicitly represents problem-solving methods,
- structural knowledge or terminology relating to the application domain, and
- support or descriptive knowledge which encodes the causal model of the application domain^{[5][6]} (see fig. 3).

Dialogue manager and problem-solver

We believe that, in order to generate versatile and adequate explanations, an expert system should consist of at least two cooperating components, which we call the problem-solver and the dialogue manager.

The *problem-solver* (*PS* from now on) works with symbols, like any computer program. These symbols

are intended to *represent* domain knowledge in the *KB*, but do not have any significance in themselves. The intended interpretation of this symbolic knowledge resides in the concepts and principles of the application domain as these are perceived by experts and knowledge engineers. In more technical terms, the *PS* and its *KB* encode a symbolic system whose *standard model* is the real world as perceived by experts and knowledge engineers.

The interface to the *PS* is a symbolic *query language*. User problems must be transformed into symbolic queries before the *PS* can solve them. Similarly, an answer from the *PS* is a string of symbols, derived from the query and the symbolic contents of the *KB*, which must be translated back into a suitable model of the application domain. So that they can put problems to the *PS* in the interface language and receive the answers in the same language, users must be familiar with the standard interpretations for symbols. This is usually the case for domain experts and knowledge engineers. Non-expert users, on the other hand, require more help with interpretation.

The *dialogue manager* (*DM* from now on) acts as a mediator in such a step. Its input function consists in translating user requests into corresponding queries that can be accepted by the *PS* (see fig. 4). Conversely, in its output function, it translates the symbolic answer produced by the *PS* into the user model of the application domain. This user model can be different from the expert's standard model: this is illustrated by the misunderstandings that occur in real life between a human expert and a non-expert.

The syntax for communicating with the user can be rudimentary, like computer menus, or it can take a more complicated form, like a manageable subset of a natural language. For example, the output function of the *DM* may be no more than generating natural-language paraphrases of *PS* answers from pre-designed ('canned') natural language templates^[3]. But comprehensive support for natural language interaction not only requires that the *DM* shall understand the literal meaning of user utterances; it also requires that it shall recognize user intentions and relate them to system capabilities and to system expectations about the way in which users formulate problems. This calls for advanced techniques and knowledge structures for dialogue management^[7], such as planning and monitoring the progress of dialogues or understanding what the user means (e.g. user utterances that imply requests without actually saying so). To eliminate mismatch between the experts' model and the users' intuition, a good *DM* must also update data about the users themselves, reflecting their current familiarity with the domain, and the *DM* must

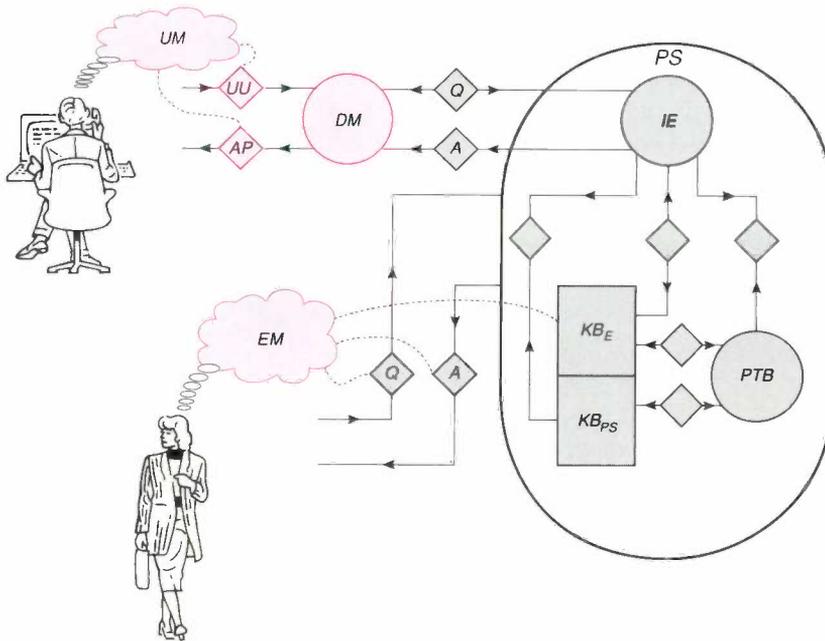


Fig. 4. A general architecture for explanation generation. Expert systems should consist of at least two cooperating parts, a problem-solver PS and a dialogue manager DM . The problem-solver is addressed by a user through the dialogue manager DM , an intermediary that translates user utterances UU into symbolic answers AP in the user languages (e.g. natural language). The DM accepts inputs and formulates answers that can be understood by the user within his view UM of the information contained in the knowledge base. Specialists (application experts and knowledge engineers) access the PS directly, however. The knowledge base KB is a direct translation of their view EM of the application domain.

choose, for each user and in each situation, the most appropriate way of referring to domain concepts^[8]. The more naive the non-expert user, the larger the discrepancy between the experts' model and the user's model is likely to be, and the more important the role of the DM . An alternative to a complex DM is a human mediator between the system and the non-expert users. For a financial counselling system, for instance, this might be a member of the bank staff.

Research on the generation of explanations is intended to improve the operation of some component or other of the architecture. Work on the DM side is concerned with improving the presentation of explanations to users. This covers a vast area with several difficult and essentially unsolved problems, like the true understanding of natural language. Work on the PS side is concerned with the identification of the knowledge that should be represented, the organization of that knowledge in the KB , and the design of strategies for generating satisfactory explanations from the KB . When both components were tackled simultaneously, there was often confusion.

Our work is focused on the PS part of the architecture. Our objective is to define a formalism with tools to represent several types of knowledge that will give meaningful answers to queries and explanations.

Generation of explanations by the problem-solver

A logical framework for explanations

Logic as specification tool. To provide a sound framework for explanations, reasoning based on the contents of the KB must be possible, and it must also be possible to specify the relationship between an answer produced by the PS and the knowledge that was used to produce it. In practice, this reasoning is only feasible if the KB and its relationships with a PS answer are described in a precise declarative manner, i.e. without references to the execution strategy of the inference engine of the PS .

One of the most widely used and best understood frameworks for formulating precise declarative specifications is *formal logic* (in practice, the first-order predicate calculus and its extensions). First-order logic has long been studied as a formalism for modelling

[5] W. J. Clancey, The epistemology of a rule-based expert system — a framework for explanation, *Artif. Intell.* 20, 215-251, 1983.

[6] W. Swartout and S. Smoliar, On making expert systems more like experts, *Expert Syst.* 4, 196-207, 1987.

[7] P. Decitre, T. Grossi, C. Jullien and J.-P. Solvay, Planning for problem formulation in advice-giving dialogue, *Proc. 3rd Conf. of the Eur. Chapter of the Association for Computational Linguistics, Copenhagen 1987*, pp. 186-190.

[8] R. Cole and J. Jerrams-Smith, Structuring knowledge for tailored explanations, *Proc. 4th Alvey Workshop on Explanations, Manchester 1988*.

sound reasoning. It is therefore at least natural (and we think it is essential) that the generation of explanations should be grounded in such a framework, in which the relationships between a request and its answer can be precisely stated.

In principle, this can be done by associating every *KB*, irrespective of its implementation formalism, with a *logical theory* that is its declarative reading (see *Table I*). The *axioms* of the theory express a declarative interpretation of the rules of the *KB*. The evaluation of a query by the action of the inference engine on the *KB* is reflected, in the logical theory, as a *formal deduction* from the axioms. The answer computed for the query is then interpreted as a *theorem* of the theory. The explanations or justifications of the validity of the answer can be derived from a *proof tree* produced by the formal deduction, that is, a structured record of the *inference rules* applied to constitute a proof.

Logic-based versus rule-based formalism as implementation tool. The interpretation of the execution of the *PS* inference engine as a formal deduction process requires that it should be possible to describe this inference engine in terms of the laws of logic, which is not always the case.

The computer language PROLOG^[9], for instance, does provide such compatibility, at least when restricted to its logical features. The rules of a *KB* written in PROLOG have a direct declarative reading as logical implications of the form *IF premise THEN conclusion*. The *premise* part consists of a logical formula containing logical connectives like AND, OR, NOT; the *conclusion* is an atomic formula (i.e. an indivisible formula). Examples of simple rules are given below (small print). The PROLOG inference engine can be interpreted as making deductions from the theory associated with the *KB*^[10].

When PROLOG is restricted to its logical features, rules have a direct declarative reading. Examples of simple rules are:

$$p(X) \text{ IF } t(X),$$

$$p(X) \text{ IF } s(X).$$

The first rule, for example, is as follows: if the premise $t(X)$ is true, then so is the conclusion $p(X)$ (with the usual notation for first-order logic, this is the implication: $\forall X t(X) \supset p(X)$, i.e. each value of X that makes t true also makes p true).

In contrast, many expert systems are implemented in formalisms derived from the early *rule-based systems*. In such systems the interpretation of a rule is more operational (*IF conditions THEN action*): roughly, when all *conditions* in a rule are satisfied with data derived from the query and the contents of the

KB, then the *action* part is executed by the inference engine. The action usually contributes to enabling conditions of other rules. The order in which rules are executed by the inference engine, and therefore the order in which the actions are performed, is often so bizarre that it is difficult to characterize the *KB* logically. For a *KB* of some complexity, the definition of

Table I. A knowledge base as a logical theory

Knowledge base	Logical theory
Rule	Axiom
Query evaluation	Formal deduction
Answer	Theorem
Evaluation trace	Proof tree

rules becomes cluttered up with programming tricks that enforce an appropriate sequence of rule execution with the inadequate strategy of the inference engine. These programming tricks also obscure the logical meaning of the *KB*^[11] and sound reasoning about its contents or sound justification of an answer from the *PS* becomes practically impossible.

If the strictest attention is paid to detail, it is not necessarily impossible to code a *KB* in a rule-based system in a manner that reflects clear declarative semantics. But if anything more than *ad hoc* explanation is contemplated, we think it would be more convenient to use a logic-based framework.

Trace-based explanations

Even when it is feasible in practice to construct a logical version of the *KB*, the answers from the *PS*, and their relationships, we are still left with the issue of just which inference system and logical theory to choose for expressing them. In particular, the choice of the *inference system* (i.e. the set of inference rules available for performing formal deductions) influences the shape of the proof tree, which records the inference rules applied to establish that a query is a theorem of the theory associated with the *KB*.

Most procedures for logic-based problem-solving (including PROLOG) are based on a version of the *resolution principle*. Resolution is a powerful, general, and efficient inference rule. Therefore conventional logic-based systems^{[4][12][13]} derive their how-, why-, and why-not-explanations from a trace of the evaluation process that records all the resolution steps performed by the inference engine during the evaluation of the query.

However, there are problems with the direct use of resolution as a basis for the generation of explanations (we also refer to the *evaluation tree* or *trace* as the deduction trace with resolution to contrast it with the *proof tree* on which explanations are based).

First, resolution is not very suitable for the production of human-readable proofs. For producing clearer proof trees, a 'natural deduction' inference system is better: it consists of clearer inference rules, similar to classical syllogisms, and therefore closer to human reasoning, with each inference and each rule more specific than resolution, so that more rules are required.

Another problem is to be found in the explanation of failure, i.e. in the strategy for why-not-explanations. Fig. 5 shows the evaluation tree formed by the inference engine of PROLOG when executing the query $p(a)$ on the simple KB consisting only of the rules

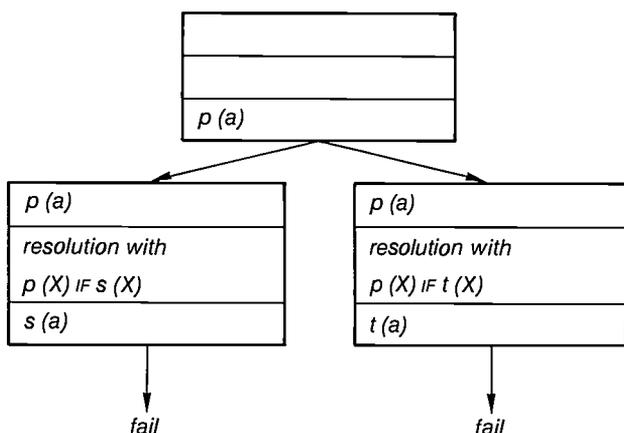


Fig. 5. Evaluation tree for $p(a)$. Each 3-part node records an inference step (with resolution) by giving: (1) a goal (formula whose verification is attempted); (2) the inference rule applied (resolution with a KB rule); (3) the result of performing the inference step (one or more formulae to be verified). For example, from $p(a)$ and the rule $p(X) \text{ IF } s(X)$, resolution infers $s(a)$. The figure shows two attempts to verify $p(a)$ that failed successively (the branching in the tree is an OR-branching), after noticing that $s(a)$ (or $t(a)$), is present as a fact in the KB and that no rule of the KB applies to $s(a)$ (or $t(a)$).

mentioned above. All attempts to establish the query $p(a)$ end in failure, as shown in fig. 5. So, in this conventional logic-based-system approach, the why-not-explanation reviews the branches of the tree in turn, which all led to dead ends (failure branches):

1. an attempt to establish $p(a)$ failed because $s(a)$ failed.
2. an attempt to establish $p(a)$ failed because $t(a)$ failed.

With the given KB, there is no other possibility for verifying the query $p(a)$, so this situation can indeed be interpreted as justifying the failure of the query.

Thus, while how-explanations justify an answer to a successful query by producing the branch of the evaluation tree that led to the answer, why-not-explanations have to justify that all attempts to produce a solution failed. In this conventional approach, how-explanations are essentially equivalent to a proof, while why-not-explanations are not, since in essence they list logical sentences which, if true, would lead to the production of an answer. Any single failure branch accounts for only a part of the justification of the occurrence of a failure. Nevertheless, with a more powerful logical theory for modelling negation, the failure to verify a query ($p(a)$ in our example) as a theorem of the theory associated with the KB would be equivalent to the fact that the negation of the query $p(a)$ is a theorem in that theory. The proof tree of the negated query then contains all the information necessary to justify the failure of the original query.

We therefore designed an approach of this nature that generates actual proofs for both how- and why-not-explanations [14]. Our inference system converts any finite evaluation tree of the PROLOG inference engine into an equivalent natural deduction proof in a theory defining the logical contents of the KB. We also designed the corresponding interpreter to form a proof tree for any query. Successes and failures of query evaluation can then be treated more symmetrically, since both can be justified by a proof (tree) as theorems of the theory.

For our system, a proof tree giving a general account of the failure of the example query $p(a)$ looks like the diagram in fig. 6. Each node records an inference rule that has been applied to prove that the negation of the query $\text{NOT } p(a)$ is a theorem. An actual proof of the query then states: $p(a)$ is FALSE because ($s(a)$ OR $t(a)$) is FALSE, that is, because both $s(a)$ and $t(a)$ are FALSE.

Decorating proof trees with extra-logical information

Proof trees may also be refined and 'decorated' with extra-logical information to provide more de-

[9] L. Sterling and E. Shapiro, The art of Prolog, Logic Programming Series, MIT Press, Cambridge 1986.

[10] J. Lloyd, Foundations of logic programming, 2nd edition, Springer, Berlin 1987.

[11] A. Bruffaerts, E. Henin and V. Marlair, Modelling and encoding knowledge for expert systems, Deliverable 8, ESPRIT Project ESTEAM, 1986.

[12] L. Sterling and M. Lalee, An explanation shell for expert systems, Comput. Intell. 2, 136-141, 1986.

[13] A. Walker, M. McCord, J. F. Sowa and W. G. Wilson, Knowledge systems and Prolog, Addison-Wesley, Reading 1987.

[14] We use 'Clark's completed database' here as the logical theory associated with the database. K. L. Clark, Negation as failure, in: Logic and databases, H. Gallaire and J. Minker (eds), Plenum, New York 1978, pp. 293-322; A. Bruffaerts and E. Henin, Proof trees for negation as failure or yet another Prolog meta-interpreter, Proc. 5th Int. Conf. Symp. on Logic programming, Seattle, Wash., 1988, pp. 343-358.

tailed input to the *DM*. We have described^[15] the definition and the implementation of a formalism that marks the premises of a rule as either contextual, essential, or auxiliary, to distinguish their specific role. The formalism also handles sets and parametrized formulae while retaining the possibility of generating proof trees. Sets and parametrized formulae are convenient for formulating constraints, i.e. logical formulae that state properties of the solution and are constructed incrementally as problem-solving proceeds.

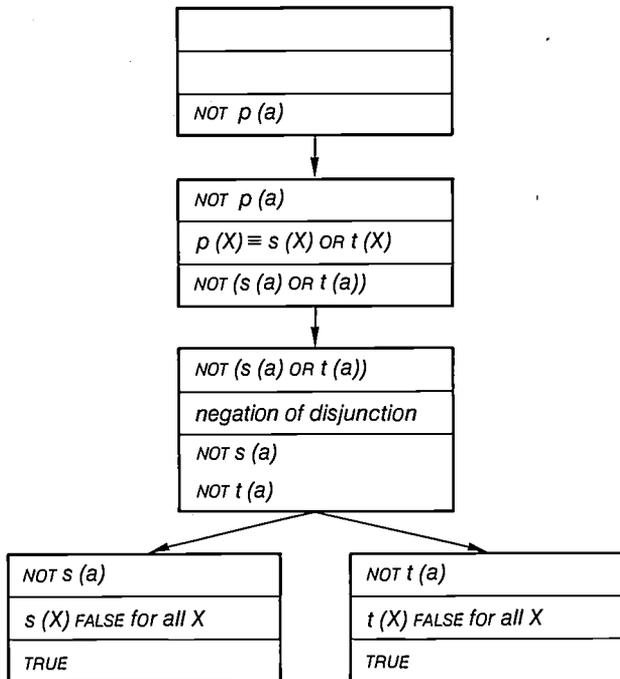


Fig. 6. A proof tree for $\text{NOT } p(a)$. The inference steps are based on a natural deduction system. There is really only one proof (one branch) in the tree (the branching at the bottom is an AND-branching): to report success itself, the parent node waits for both $\text{NOT } s(a)$ and $\text{NOT } t(a)$ to report success.

With a uniform representation of problem-solving knowledge and explanatory knowledge, the formation of proof trees becomes a powerful and general mechanism for exploiting that knowledge. Indeed, the same technique can be used for tracing both the computation of a solution by the *PS* and the subsequent application of explanatory knowledge to that solution.

More conceptual explanations

As argued above in the section on 'Early attempts at generating explanations', it is also interesting to cater for explanations not directly based on the *PS* proof tree of some problem-solving query. These require dedicated domain knowledge, which may re-

late to the terminology, the conceptual structure, or the causal model of the domain. For it to be available, such problem knowledge has to be encoded into the *KB*, and it can then be directly queried, just like problem-solving knowledge. The border between knowledge for problem-solving and knowledge for explanation generation then becomes fuzzy and, moreover, it is not essential to try to distinguish between them. Generating more conceptual explanations becomes itself a type of problem to be solved, and quite naturally it requires dedicated knowledge.

Explanation knowledge is usually meta-knowledge, i.e. knowledge about knowledge. Therefore, entities that encode meta-knowledge must be allowed to refer to other entities of the same nature that encode knowledge. This calls for a formalism where all knowledge-encoding entities are treated alike.

We are developing a logic-based object-oriented formalism^[16] which allows uniform encoding of knowledge and meta-knowledge and the linking of entities from several conceptual levels in the *KB*. The formalism is based on a powerful first-order logical theory in which individual objects, classes of objects, attributes of objects, and associations among objects are in turn all treated as objects, i.e. as elements of a 'global universe of discourse'. This means that the normal structuring principles of the object paradigm (see box) are integrated into logic. This homogeneity of the object theory may be compared with that of set theory, where functions and relations are ordinary sets.

The formalism permits a variety of conceptual explanations to be generated, provided, of course, that the corresponding knowledge is present in the *KB*. Because of the uniform representation of objects, all items of knowledge can be represented in the *KB* and queried directly through the *PS* interface.

The justification of a solution produced by the *PS* is provided as the proof tree obtained from another query to the *PS* requesting verification that the solution complies with the fundamental principles of the domain.

For example, if explanations are requested about a concept of the domain (normally represented as a class in the object formalism (see *fig. 7*), information is extracted from the network of objects, starting from the representation of the concept requested, exploring its attributes (the immediate properties of the concept), its associations with other concepts, its immediate superclasses (the closest generalizations of the concept), their attributes and their associations, and so on until a suitable level of detail is obtained. This exploration is conducted by appropriate queries to the *KB* through the *PS* interface.

Object-oriented approach to knowledge representation

Object orientation is more an approach than a specific set of constructs.

As an informal definition, an object-oriented formalism allows the representation of every important entity in an application domain as one object in the formalism.

A class is an abstraction of a collection of objects with similar properties.

In an object-oriented approach, a *KB* is structured around a hierarchy of classes. A class can be connected with classes of lower rank (subclasses) and with classes of higher rank (superclasses).

A class inherits properties (or attributes) of its superclasses while defining some properties of its own.

Objects are instances of at least one class and of all its superclasses.

General structuring mechanisms provide a powerful modelling methodology for:

- the relationships between an object and its class and superclasses (instantiation-classification)
- the relationships between a class and its superclasses in the hierarchy (specialization-generalization)
- the representation of associations among objects as other objects (decomposition-aggregation).

In another example, if a user wants to know the differences between two similar concepts, their closest common superclass will be inspected (the closest concept that is a common generalization), their common attributes (if any), their specific attributes, and so on.

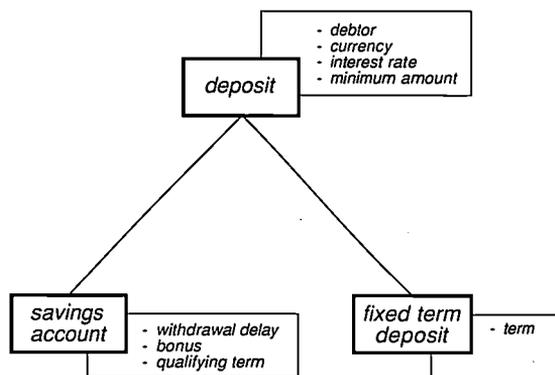


Fig. 7. Two classes with one common superclass and some of their attributes. In the financial domain, savings accounts and fixed-term deposits are special cases of deposits. Deposits are represented as a superclass of the classes of savings accounts and fixed-term deposits. The attributes of savings accounts, for example, comprise the attributes specifically associated with the class (notice required, etc.) plus those of the superclass (debtor etc.).

[15] A. Bruffaerts, E. Henin and V. Marlair, How-explanations for real-life expert systems, Deliverable 15, ESPRIT Project ESTEAM, 1987.

[16] A. Bruffaerts and E. Henin, A logic-based object-oriented knowledge representation language, Deliverable 6 - Part 2, ESPRIT Project ESTEAM-316, 1988.

Summary. In this article we have argued that a wide spectrum of explanations can be useful in an expert system. We have looked briefly at attempts to generate some of these explanations. A general architecture has been proposed for expert systems, clearly distinguishing between the functions of two cooperating agents, a dialogue manager and a problem-solver. We have argued that a logical framework is most convenient for a precise declarative version of the knowledge base and of its relationships to query answers from the problem-solver. Our own work is aimed at the development of a powerful problem-solver component along those lines. Formalisms and tools have been developed for generating a form of proof tree that offers a sound basis for the conventional trace-based how-, why-, why-not-explanations. A logic-based object-oriented knowledge-representation formalism is being designed to support a uniform encoding of knowledge and meta-knowledge for the generation of more conceptual explanations.

HS3: an advanced bipolar-IC technology

G. Conner and R. H. Lane

There are three widely accepted fundamentals of integrated-circuit technology: cost per function decreases, performance increases and the applications expand. They necessitated drastic reduction in bipolar-transistor dimensions. This has stimulated the use of increasingly sophisticated process tools: plasma etching, ion implantation, low-pressure film deposition and 'near-one-micron' photolithography. The new bipolar-IC manufacturing process HS3 from the Philips Research Laboratories at Sunnyvale, U.S.A., is an example of merging these complex technologies in a single cost-effective process.

Introduction

Bipolar transistors are used in analog circuits, in mixed analog/digital circuits and in high-speed digital circuits. They derive their name from the fact that both electrons and holes contribute to the conduction. In MOS transistors (MOS: metal-oxide semiconductor), on the other hand, the conduction is due to the transport of holes *or* electrons: there is only one type of charge carrier^[1]. Bipolar integrated circuits are built up from npn transistors and in some cases pnp transistors.

As is so often the case, development of the Signetics HS3 process for the manufacture of bipolar ICs was started as a response to both internal needs and external competition. Technologies for bipolar ICs with SiO₂ isolation were widely used and many companies were working on second-generation versions. Signetics, eager to maintain a competitive position in high-speed circuits, also pushed for improved performance of their existing HS2 manufacturing-process family. The HS2 process had been developed previously at Philips Research Laboratories, Eindhoven.

To be successful, HS3 would need twice the speed and density of the HS2 process family. That realization required the application of many new technol-

ogies. Some of them were used for the first time on a bipolar-IC production process. The major technological advances included the use of direct-stepper photolithography for all masking layers, r.f. plasma-reactive etching of most thin films, isolation with planar-oxide LOCOS^[2], ion implantation of all dopants, and statistical process control. It soon became clear that HS3 would be a vehicle for introducing a wide range of new concepts into the bipolar-IC production environment. The new bipolar-IC process that has been developed at the Philips Signetics Research Laboratories can be characterized as an all-implanted, planar-oxide-isolated bipolar-IC process with multi-level metallization.

In the rest of the article we shall take a closer look at the manufacturing process. The devices that can be fabricated on a silicon wafer will then be discussed. Finally we shall mention some applications.

The manufacturing process

General

An IC-manufacturing process can be assessed to a first order by the number of photolithographic steps or 'masking steps' required to manufacture the finished product. In each masking step a photosensitive

G. Conner and R. H. Lane are with Philips Research Laboratories, Sunnyvale, USA.

resist layer on the silicon wafer is exposed. After development of the resist the exposed areas — in the case of a 'positive' photoresist — or the unexposed areas — in the case of a 'negative' photoresist — are subjected to a processing step, such as etching or ion implantation. The process complexity can easily be estimated using the number of masking steps required as a 'figure of merit'.

The number of masking steps required to manufacture a single type of bipolar integrated circuit is limited by designing the HS3 process as a combination of a 'core' process with a number of 'modular additions'. Each modular addition can be considered as an option to manufacture a specific kind of bipolar IC. Each option requires one additional masking step, see *fig. 1*.

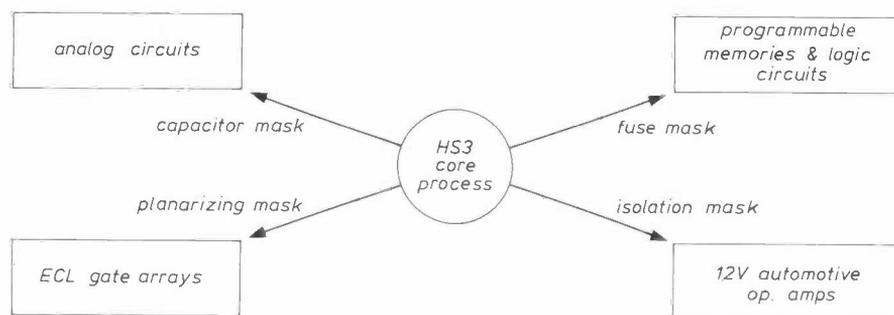


Fig. 1. The HS3 core process and its optional additions, each requiring an extra masking step.

This modular concept 'designs flexibility into' the process. While no single process could be expected to fulfill all needs, the modular process family HS3 can be used to manufacture quite a range of bipolar integrated circuits:

- analog circuits,
- gate arrays for ECL circuits (ECL: emitter-coupled logic),
- programmable memories and logic circuits and
- operational amplifiers for automotive applications with 12-V supply.

To the circuit designer the concept of *fig. 1* has many advantages. The same set of transistor models for circuit simulation can be used in many different applications. In the construction of devices a single basic set of layout rules in the core process can lead to the generation of cell libraries for a variety of options. When more designers from various disciplines use the same process, a large 'knowledge base' develops with an increased sophistication in design.

Manufacturing also benefits directly from the core-process structure of *fig. 1*. As a result of the large numbers of wafers following the same process steps

the reliability of statistical process control methods improves. Fewer processes in the IC factory bring about a higher degree of focus on problems. Less equipment improves both throughput time and yield. Moreover, circuits with only small-volume requirements will be more cost-effective if run on a high-volume process.

The core process

Fig. 2 shows cross-sectional diagrams of the different HS3 process steps in manufacturing the basic npn transistor. The substrate is p material, silicon with a surplus of free holes, with a (111) surface. *Fig. 2a* shows that the process starts off with the diffusion of a buried n⁺ layer (i.e. with an extra surplus of free elec-

trons) into the substrate. This layer is an antimony-implantation and functions as the buried collector for the transistor. The layer, however, is shallow and designed to have a minimum of diffusion in the upward direction during subsequent processing.

A buried p⁺ layer is also formed (not shown). This is typically used as the lower part of the vertical 'contact hole' to connect the substrate with the metal interconnecting layers at the surface. Also, this p⁺ layer can form a low-resistance grounding grid. This will allow the designer to reduce any sensitivity to injected substrate charge, which could cause an SCR-type latch-up: short-circuiting caused by the formation of a parasitic 'silicon-controlled rectifier' via the substrate.

Fig. 2b shows how in the next process step a thin layer of arsenic-doped silicon is epitaxially grown to

[1] W. G. Gelling and F. Valster, The new centre for submicron IC technology, Philips Tech. Rev. 42, 266-273, 1985/86.

[2] J. A. Appels, E. Kooi, M. M. Paffen, J. J. H. Schatorjé and W. H. C. G. Verkuylen, Local oxidation of silicon and its application in semiconductor-device technology, Philips Res. Rep. 25, 118-132, 1970; J. A. Appels, H. Kalter and E. Kooi, Some problems of MOS technology, Philips Tech. Rev. 31, 225-236, 1970.

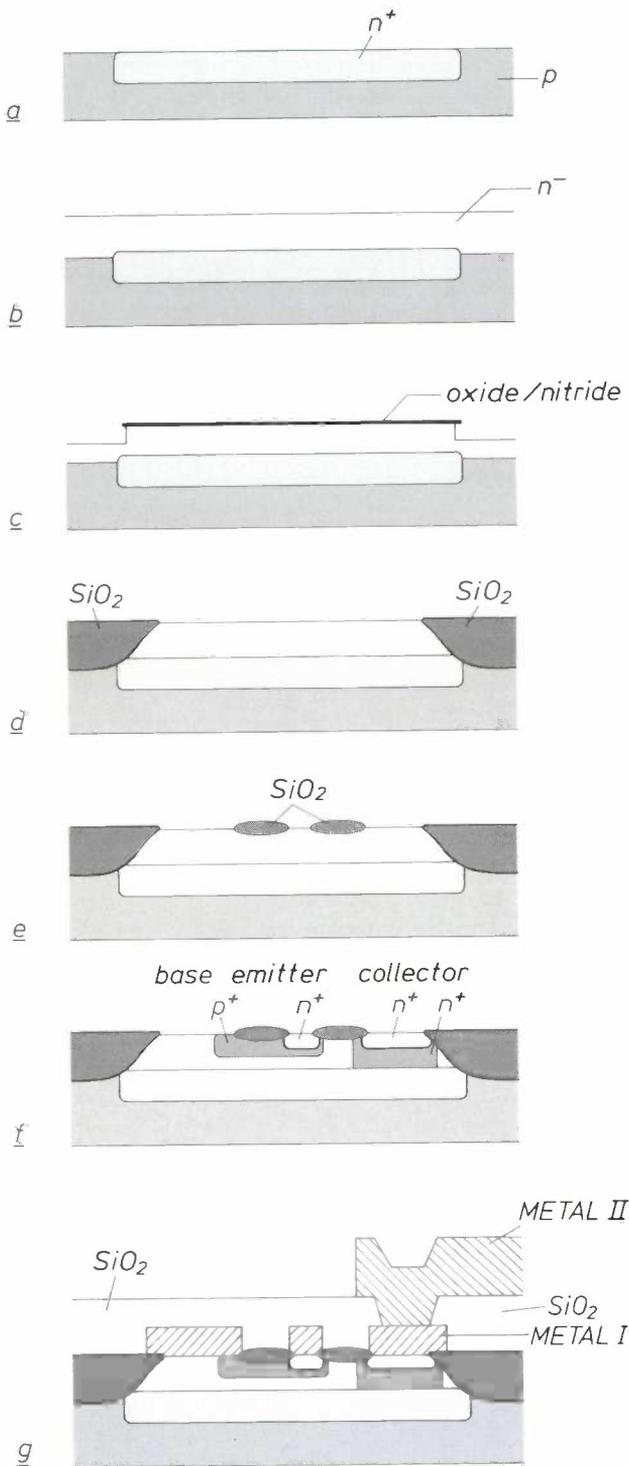


Fig. 2. The basic steps in the HS3 core process. *a)* Diffusion of an n^+ layer into a p -substrate. *b)* Epitaxial growth of an n^- layer. *c)* Alternate growth and deposition of silicon dioxide and nitride layers. These layers and half of the n^- layer are partially removed at the areas where isolating LOCOS oxide is required^[2]. *d)* Growth of a thick oxide layer in the exposed areas and planarization of this layer. *e)* Growth of a second LOCOS oxide layer. *f)* Ion implantation of collector connections (n^+), base (p^+) and emitter (n^+), partially through the second LOCOS oxide layer. *g)* Metallization: deposition of a first interconnection layer of aluminum-copper alloy, a layer of 'doped glass' and a second interconnection layer of aluminum-copper.

form the basis for the active and passive devices. The characteristics of this epitaxial layer are crucial to the process as it has a major impact on such device parameters as capacitance, breakdown voltage, and current density. The quality of this layer, measured by the number of crystal defects, has also a major contribution to the yield of accepted ICs.

At this point in the process, the isolation of individual devices is formed by using a recessed-LOCOS technique^[2]. Thin layers of silicon dioxide and nitride are alternately grown and deposited over the epitaxial surface. The area between devices to be formed later is etched in a plasma reactor: the nitride, dioxide, and about half of the epitaxial layer are removed, see fig. 2*c*.

A very thick layer of silicon dioxide is then grown in the exposed areas only. This oxide consumes the remaining epitaxial layer at these areas and replaces it as far as the previous surface. If thus separates the masked device areas physically and electrically. The silicon-dioxide surface must then be 'planarized'. First, a sacrificial layer of borophosphosilicate-glass is thermally induced to 'flow' over the wafer surface. Very selective etching then results in the surface shown in fig. 2*d*.

The LOCOS technique is now applied again, though not in the recessed areas, see fig. 2*e*. The remaining areas now define what will eventually become the contacts between the doped areas in the substrate and the metal interconnecting layers. The method ensures that a contact hole with smooth side walls will be produced later. Such a contact hole needs no further treatment to ensure excellent coverage by the metal interconnecting layers. The oxidation of the area between contact holes has the effect of reducing their size. This allows a 1.0- μm emitter to be formed using a 1.75- μm mask image.

The base, emitter and low-ohmic collector connections are next formed by ion implantation, see fig. 2*f*. Except for the emitter, each layer is implanted directly through the thin second-LOCOS layer. After this, a furnace heat treatment of short duration corrects for lattice damage in the implanted layers and determines their final depths.

Fig. 3 shows that the vertical dimensions of the layers thus formed are extremely small. The boron-doped base layer between emitter and collector, for example, has a thickness of less than 0.2 μm . The key to the high speeds of the circuits obtained is the accurate control of the thickness of these layers.

The wafers are now ready for the metallization steps in the process. First is the formation of Schottky diodes (the junctions between metal and the epitaxial silicon) and low-ohmic contacts, using sputter-depos-

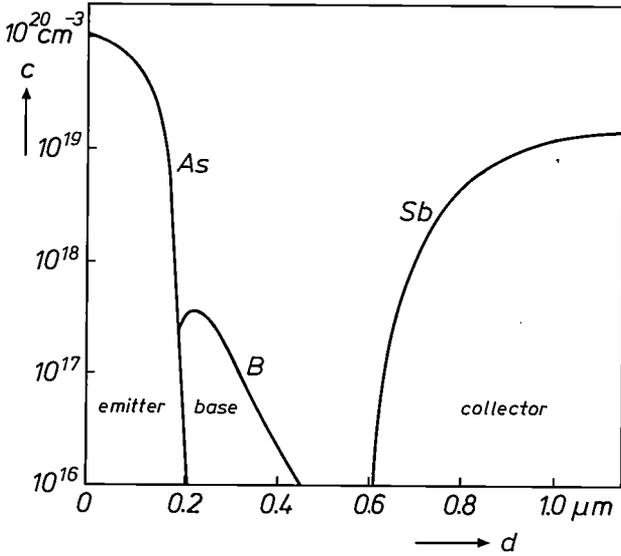


Fig. 3. Concentration profiles, concentration c (in atoms per cm^3) as a function of depth d , at the emitter.

ited platinum. The platinum is heat-treated to form platinum silicide with the silicon underneath. The silicide forms only in the open contact areas and the unreacted metal can be selectively etched. Next a barrier layer of titanium tungsten is deposited. It forms a sandwich layer between the silicide and a layer of aluminum-copper alloy, which is deposited next. After masking and plasma etching, these three layers form the first metal interconnecting layer (*METAL I* in fig. 2g).

The second and following interconnecting layers are formed by the deposition of doped silicon dioxide, plasma etching of interlayer contact holes in the oxide and addition of an aluminum-copper layer, see fig. 2g. Aluminum-copper alloy is used instead of pure aluminum to enhance the resistance of the metal film to failure due to electromigration.

All HS3 products manufactured so far have been exhaustively tested to verify the reliability of the process described above.

Devices

As with most advanced bipolar technologies, npn transistors are the main devices. To give a comparison of dimensions fig. 4 shows a plan view of typical npn transistor layouts in the former HS2 process and the recently developed HS3 process. The data in Table I reflect the performance improvement as device dimensions shrink. The current smallest npn transistor is characterized by its emitter of 1.0 by 3.0 μm . Other key parameters of this device are summarized in Table II.

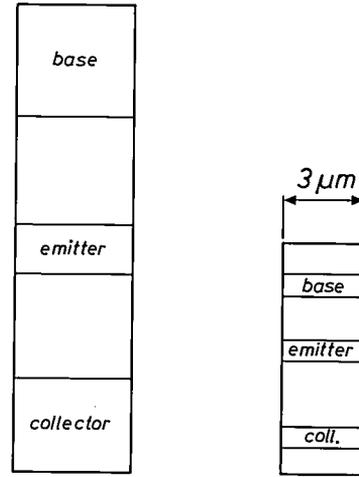


Fig. 4. Plan views of npn transistors in the HS2 process of 1982 (left) and in the HS3 process of 1987.

Table I. Parameter values that characterize the HS2 IC-manufacturing process of 1982 and the HS3 process of 1987. $f_{T,max}$ maximum cut-off frequency. C_{CB} collector-base capacity. d minimum wafer-detail size. t_1 pitch of first interconnecting layer. t_2 pitch of second interconnecting layer. t_3 pitch of third interconnecting layer (not possible in HS2). τ_G minimum gate delay at a step of 500 mV.

	$f_{T,max}$ [GHz]	C_{CB} [fF]	d [μm]	t_1 [μm]	t_2 [μm]	t_3 [μm]	τ_G [ps]
HS2	5	60	2.5	9	16	-	350
HS3	9	18	1.5	4.5	8	15	130

Table II. Parameter values that characterize an npn transistor in the HS3 process. The emitter dimensions are 1.0 by 3.0 μm . β current gain $\partial I_C / \partial I_B$. R_B base resistance. R_C collector resistance. R_E emitter resistance. BV_{CB} breakdown voltage between collector and base with open emitter. BV_{CE} breakdown voltage between collector and emitter with open base. BV_{EB} breakdown voltage between emitter and base with open collector. C_{BB} emitter-base capacitance. C_{CB} collector-base capacitance. C_{CS} collector-substrate capacitance. $f_{T,max}$ maximum cut-off frequency.

β	R_B [Ω]	R_C [Ω]	R_E [Ω]	BV_{CB} [V]	BV_{CE} [V]	BV_{EB} [V]	C_{BB} [fF]	C_{CB} [fF]	C_{CS} [fF]	$f_{T,max}$ [GHz]
120	400	70	2	15	7	5	16	18	40	9

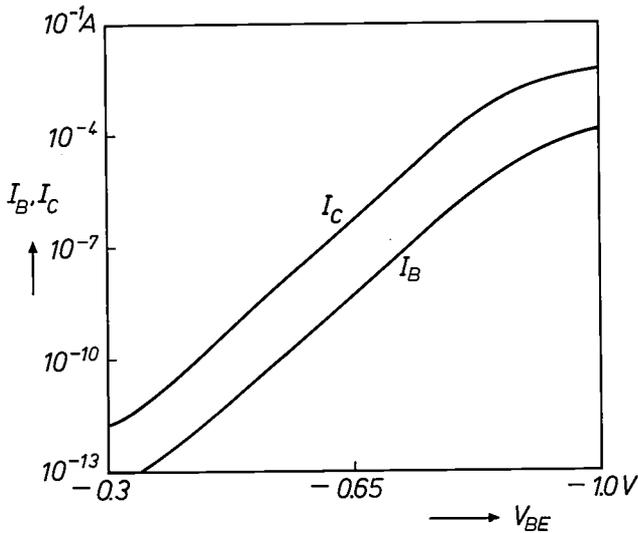


Fig. 5. Collector current I_C and base current I_B plotted against the emitter-base voltage V_{BE} . This plot is usually called a 'Gummel plot'.

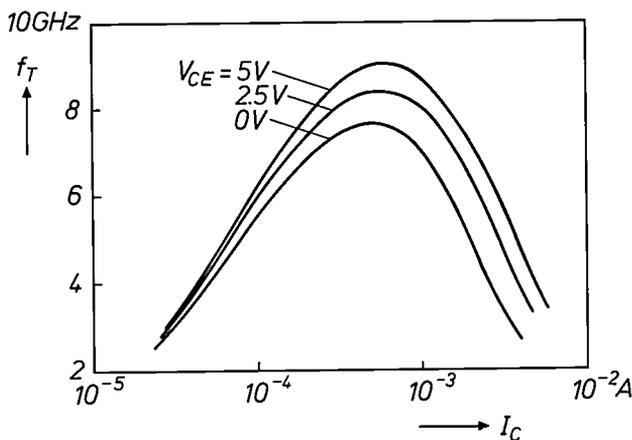


Fig. 6. Cut-off frequency f_T — unity collector-current gain frequency — as a function of collector current I_C for collector-emitter voltages V_{CE} of 0, 2.5 and 5 V. The curves demonstrate the excellent high-frequency properties of npn transistors in the HS3 process.

Two important diagrams complete our description of the npn transistor and its characteristics. Fig. 5 — usually called a 'Gummel plot' — gives the collector and base currents as a function of emitter-base voltage. It shows that the collector current approximates to an ideal exponential behaviour. Fig. 6 is a plot of the cut-off frequency f_T as a function of collector current as measured on the wafer^[3]. f_T represents the frequency at which the current gain $\beta = \partial I_C / \partial I_B$ has decreased so much that it equals unity^[4]. It is a well-known 'figure of merit' for high-speed operation.

The design of successful circuits requires many more equally complex devices, however. Among those most commonly used are: lateral and vertical pnp

transistors, high-value capacitors, resistors, programmable fuses, electrostatic-protection diodes and Schottky diodes. The HS3 process was designed with improvements to all these devices in mind. In some instances, such as the electrostatic-protection diode, new structures were invented^[5].

The electronically programmable fusing element is a good example of an HS3 option requiring one single mask (fig. 1). The normal mask for the first interconnecting layer also defines the outlines of the programmable fuses. After the AlCu metal and TiW barrier underneath have been plasma etched to define both interconnecting layer and fuses the optional fuse mask is applied. This mask allows the selective removal of AlCu in the fuse areas. This results in the formation of fuses of TiW alone. The fuses can be 'blown' by applying a somewhat higher voltage than the supply voltage across their connections.

Applications

The HS3 process has been applied to a wide variety of products. A general cross-section of the circuits designed in HS3 includes:

- an analog amplifier for a maximum frequency of 600 MHz,
- a logic circuit with TTL (transistor-transistor logic) gates,
- an analog 150-MHz preamplifier for fiber optics,
- a 50-MHz postamplifier for the same application,
- a programmable memory for ECL with an access time of 5 ns,
- ASICs (application-specific integrated circuits) for TTL for chip areas of 10 to 25 mm²,
- 12-bit analog/digital converters trimmed by programmable fuses,
- an ASIC with gate arrays for a maximum frequency of 600 MHz.

These products amply demonstrate the ability of HS3 to achieve high yields across a very wide range of chip area.

Another advantage of HS3 is the ability to mix technologies on the same chip. The best example of this is the digital/analog converter for video applications pictured in fig. 7. This circuit combines a 3-kbit high-speed digital RAM (random-access memory) with three linear digital/analog converters. The circuit can operate at 200 MHz because of the excellent performance of each section.

A further example of mixing technologies is a motor-driver interface developed for automotive applications, see fig. 8. As well as analog and digital circuits, the interface circuit also includes current drivers with a break-down voltage of more than 40 V. This inte-

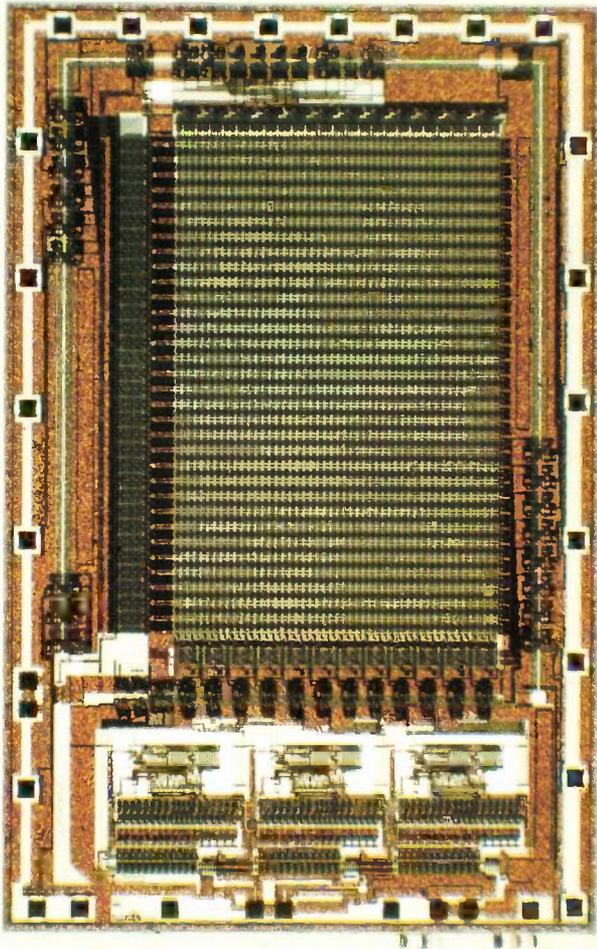


Fig. 7. A circuit for video applications that combines three digital/analog converters and a random-access memory. It can operate at 200 MHz.

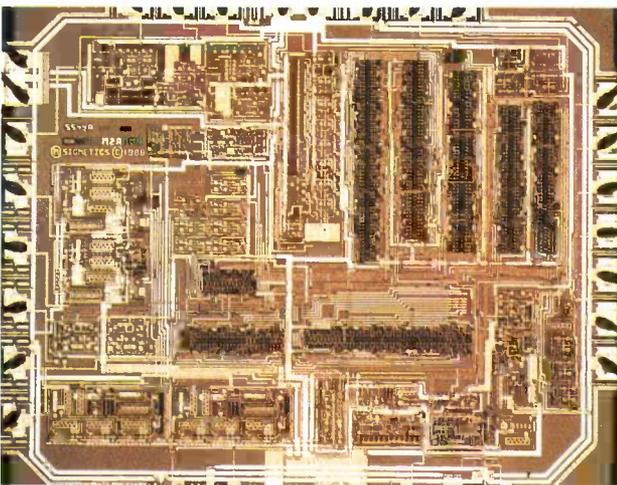


Fig. 8. A motor-driver interface circuit for automotive applications. The dense circuit combines digital, analog and high-voltage devices.

grated circuit shows that the designer has complete flexibility in selecting devices that optimize speed, voltage, packing density or analog control.

A final example of application of HS3 technology is the gate-array circuit already mentioned, see fig. 9. At a chip area of over 80 mm², this is the largest bipolar circuit ever made at Signetics. It is shown in a 148-pin package representing the final product form.

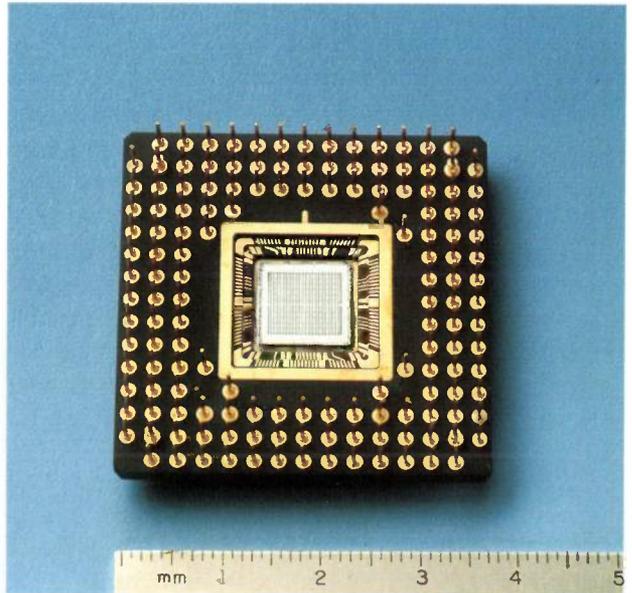


Fig. 9. A gate-array circuit in HS3 technology in a 148-pin package. The chip area is more than 80 mm².

- [3] P. J. van Wijnen, H. R. Claessen and E. A. Wolsheimer, A new straightforward calibration and correction procedure for 'on wafer' high frequency S-parameter measurements (45 MHz – 18 GHz), Proc. IEEE Bipolar Circuits and Technology Meeting, Minneapolis, Minn., 1987, pp. 70-73.
- [4] S. M. Sze, Physics of semiconductor devices, Wiley, New York 1981.
- [5] W. D. Mack and R. H. Lane, Protection device utilizing one or more subsurface diodes and associated method of manufacture, U.S. Patent No. 4736271 (5th April 1988).

Summary. HS3 has been shown to be a versatile, high-performance IC-manufacturing process. The process consists of a single core process and a number of modular additions. Each addition requires an extra masking step and corresponds to an optional product family, such as analog circuits, ECL gate arrays, memory and logic circuits, and operational amplifiers for automotive applications. High production numbers in the core process facilitate statistical process control. The excellent properties of npn transistors with 1 by 3 μm emitter are demonstrated by a Gummel plot and a plot of cut-off frequency f_T against collector current.

Research on monolithic GaAs MESFET circuits at LEP

M. Rocchi

Since 1974 a succession of research teams at LEP (Laboratoires d'Electronique et de Physique Appliquée) have been successfully investigating the potentialities of monolithic GaAs circuits in high-speed digital electronics and r.f. and microwave applications. This has led to world-first demonstrations of devices such as monolithic 12-GHz front-ends for direct satellite television reception, frequency-dividers up to 10 GHz and, more recently, eight-by-eight-bit multipliers with multiplying times less than 2 ns at a power consumption of only 350 mW. The main achievements in the past fifteen years are reviewed in the article below.

Why GaAs MESFET ICs?

A GaAs MESFET (MEtal-Semiconductor Field-Effect Transistor) is a field-effect transistor in which a metal-semiconductor rectifying contact is used for the gate. A schematic cross-section and a photograph of a GaAs MESFET structure are shown in *fig. 1*. Current conduction between the source and drain takes place in a thin n-doped GaAs layer grown on a high-resistivity ('semi-insulating') GaAs substrate. The source and drain form ohmic contacts with this layer, whereas the gate metal is insulated from it by a reverse-biased Schottky-barrier diode. The depletion region beneath the gate restricts the electric current to a narrow n-channel above the substrate. The conductance of this channel is modulated by the gate voltage V_g , which affects the width of the depletion region and hence that of the n-channel.

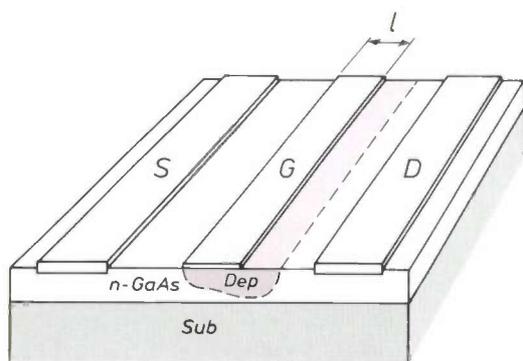
There are basically two different types of GaAs MESFET, depending on the doping and the electron density under the gate. If an n-channel exists at $V_g = 0$, a negative gate voltage must be applied to suppress the channel conductance. This type is called the normally-on (depletion-type) n-channel MESFET. If the channel conductance at $V_g = 0$ is very low, a positive gate voltage is required to form the n-

channel. This type is called the normally-off (enhancement-type) n-channel MESFET.

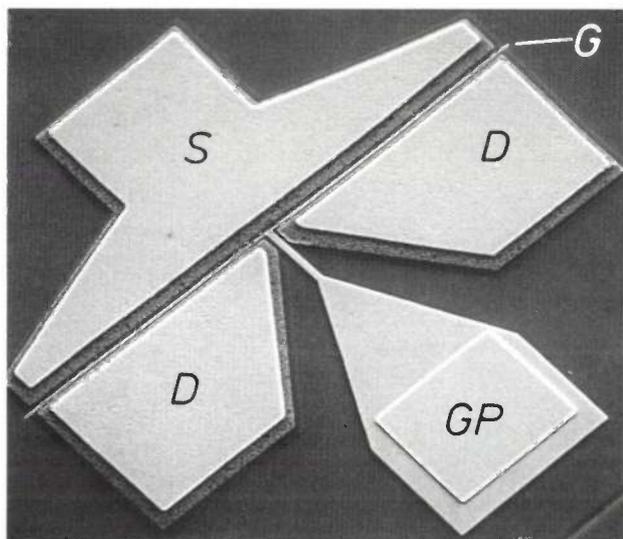
After the pioneering work of Dr Jutzi's team at IBM in Zürich on GaAs normally-off static random-access memories (SRAMs)^[1] the ingredients for high-frequency circuits became available. The first was the n-channel GaAs MESFET with its excellent high-frequency performance, especially when the gate length is reduced to submicron dimensions. The second ingredient was a semi-insulating GaAs substrate (with a resistivity of about $10^8 \Omega\text{cm}$) for simple active-device insulation, with low interwiring capacitances and a high resistance to damaging radiation such as fast neutrons or X-rays.

The outstanding high-frequency performance of n-channel MESFETs is due to the excellent transport properties of conduction electrons in active GaAs layers: a high mobility at low electric field ($\geq 4000 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ at a doping level of $2 \times 10^{17} \text{ cm}^{-3}$) and a high saturated electron velocity ($2.5 \times 10^7 \text{ cm/s}$) with marked overshoot and quasi-ballistic effects in structures with a gate length shorter than $0.5 \mu\text{m}$. As a direct consequence, the electron transit time τ beneath the Schottky gate can be as short as 2.6 ps for a gate length of only $0.2 \mu\text{m}$. The frequencies at which the transistor still shows some current gain can therefore be very high, with a unity-current-gain cut-off frequency $f_T (= 1/2\pi\tau)$ as high as 60 GHz. Also, the

Dr M. Rocchi, formerly with Laboratoires d'Electronique et de Physique Appliquée (LEP), Limeil-Brévannes, France, is now with Philips Research Laboratories, Eindhoven. Many colleagues at LEP have contributed to the work described here, which is still being actively pursued.



a



b

Fig. 1. *a*) Schematic cross-section of a GaAs MESFET consisting of a semi-insulating GaAs substrate *Sub*, an n-doped GaAs layer, a source *S*, a metal gate *G* (of length *l*) and a drain *D*. The gate is insulated from the n-GaAs layer by a reversed-biased Schottky-barrier diode. Owing to the depletion region *Dep* beneath the gate, the current is restricted to a narrow n-channel. *b*) Photograph of a GaAs MESFET with a gate length of 0.3 μm . The gate *G*, midway between the source *S* and the drain *D*, is connected to the gate pad *GP*.

input impedance of a MESFET consists of a very low gate resistance ($R \approx 0.01 \Omega/\square$) in series with a small capacitance ($C \approx 1 \text{ fF}/\mu\text{m}^2$), resulting in a negligible input RC time constant compared with the electron transit time.

Three-port devices such as MESFETs with high f_T -values and negligible input RC constants, along with low interwiring capacitances, are the main prerequisites for high-speed digital ICs. In fact, frequency-dividers about three times faster than their silicon counterparts have been demonstrated at the Hewlett-Packard Laboratories^[2] and at LEP^[3]. This triggered off world-wide research on monolithic digital GaAs circuits, with impressive results such as

the demonstration (in 1987) of frequency-dividers based on master-slave flip-flops with a maximum operating frequency of 26 GHz.

The features of GaAs MESFETs described above yield other pay-offs such as very low noise figures at frequencies above 100 MHz and up to f_T . For instance, a typical performance of a MESFET with an ion-implanted active layer and a gate length of 0.5 μm is a noise figure of 1.6 dB at 12 GHz with an associated gain of 9 dB, which is a few dB better than obtained with advanced bipolar transistors based on silicon. GaAs MESFETs can also be used to deliver microwave power. A power-gain cut-off frequency as high as 80-120 GHz can be obtained, which is far beyond the capabilities of silicon devices. For instance, 2.4 mm \times 0.7 μm MESFETs designed and fabricated at LEP in 1984 had a power output of 1 W at 18 GHz with a linear gain of 6.6 dB.

By using such high-performance devices, 'hybrid' microwave circuits on an alumina substrate — commonly referred to as microwave integrated circuits (MICs) — have been fabricated for frequencies up to 30 GHz^[4]. However, fully monolithic circuits can also be made, since distributed lines and passive elements as well as GaAs MESFETs can be fabricated on semi-insulating GaAs substrates. The first demonstrations of monolithic microwave integrated circuits (MMICs) were reported in 1976^[5]. In some respects they offer better performance^[6]; they are smaller and should prove most appropriate for large-volume low-cost applications.

Digital GaAs ICs

First-generation devices (1974-1981)

The evolution of digital GaAs ICs based on MESFETs has always been dependent on the ease of control of the transistor threshold voltage V_T , i.e. the minimum value of the gate voltage required to turn the transistor on. This value depends on the doping of the active layer and the electron-density profile

- [1] W. Jutzi, Direct coupled circuits with normally-off GaAs-MESFET's at 4.2 °K, Arch. Elektron. Übertragungstech. (AEÜ) 25, 595-598, 1971.
- [2] R. L. van Tuyl, C. A. Liechti, R. E. Lee and E. Gowen, GaAs MESFET logic with 4-GHz clock rate, IEEE J. SC-12, 485-496, 1977.
- [3] M. Cathelin, G. Durand, M. Gavant and M. Rocchi, 5 GHz binary frequency division on GaAs, Electron. Lett. 16, 535-536, 1980.
- [4] P. Gamand, A complete small size 2 to 30 GHz hybrid distributed amplifier using a novel design technique, IEEE MTT-S Int. Microwave Symp. Digest, Baltimore, Md., 1986, pp. 343-346.
- [5] R. S. Pengelly and J. A. Turner, Monolithic broadband GaAs F.E.T. amplifiers, Electron. Lett. 12, 251-252, 1976.
- [6] M. Levent-Villegas, M. Soulard and C. Villalon, S-band GaAs monolithic linear dual phase modulator, Proc. 17th Eur. Microwave Conf., Rome 1987, pp. 421-426.

beneath the gate. Early processes were based on the use of epitaxial active layers since ion-implanted GaAs layers could not be annealed reproducibly until 1980. Epitaxial layers might have a very high electron mobility, but they were not very uniform. Consequently, only normally-on digital circuits using MESFETs with negative V_T were then considered.

First-generation digital ICs developed at LEP were based on aluminium-gate MESFETs with gate lengths ranging from 0.6 to 1 μm and a V_T -value of -2.5 V . Submicron gate lengths were achieved by carefully controlling the underetching of the aluminium-gate strip beneath a photoresist line. To insulate the transistors electrically, a non-annealed boron implant was developed, resulting in a fully planar process.

A typical logic-gate configuration built up from such transistors consists of a logic block similar to that of silicon NMOS logic gates (NMOS stands for N-channel Metal Oxide Semiconductor). Since the input transistors are of the normally-on type, the voltage levels at the output of the logic block are not compatible with the input levels of the next logic block. A voltage-buffer stage is then required to shift these levels downwards to achieve the desired input/output compatibility. Such buffered FET logic (BFL) gates have a rather high power consumption since the relatively large negative value of V_T necessitates the use of two supply voltages. The buffer stage consists of a common drain transistor, two level-shifting diodes and a current-source MESFET to set the current and therefore the voltage drop across the diodes. The low logic level should be close to $V_T = -2\text{ V}$ and consequently the negative supply voltage is -3 V (with a drop of about 1 V across the current source). The high logic level is close to 0.5 V and a positive supply voltage of $3\text{ to }4\text{ V}$ is required. The switching speed of these logic gates is very good, however. Typical performance values are a propagation delay time of 95 ps at a power consumption of 20 mW per gate and a fan-out (number of inputs connected to one output) of 2. Recently it has been shown that the same speed could be achieved with a power consumption of only 2 mW per gate when using MESFETs with a less negative V_T (-0.5 V) [7].

The first demonstration of small-scale integration (SSI) using these BFL gates was in static frequency-dividers in the form of divide-by-two and divide-by-eight circuits. Since these circuits require complementary inputs, these complementary signals were generated directly on the chip. A maximum operating frequency as high as 5.3 GHz was observed. For operation at higher frequencies a dynamic flip-flop divider-by-two was designed. Careful optimization of the pass transistors was required since the gate current

rises rapidly when the input diode is forward-biased. This pioneering work resulted in a divide-by-two circuit operating at 10.5 GHz with a power consumption of 160 mW [8]. More recently, researchers at Hughes Laboratories have demonstrated dynamic frequency-dividers at 26 GHz based on the same design with normally-on MESFETs with a gate length of only $0.2\text{ }\mu\text{m}$. These were fabricated with a Philips electron-beam pattern generator.

Second-generation devices (1981-1989)

When ion implantation became a reliable technique for producing GaAs ICs, normally-off devices (with $V_T \geq 0\text{ V}$) were immediately investigated since they offered promise of high packing density, and — very important — low power consumption ($\leq 1\text{ mW}$ per

Table I. Applications, examples of functions and performance (including output buffers) of some of the digital GaAs ICs designed and fabricated at LEP.

Applications	Examples of functions	Performance
Instrumentation systems	Divider-by-2	Maximum frequency = 11 GHz Power consumption = 140 mW
	Divider-by-8	Maximum frequency = 6 GHz Power consumption = 20 mW
	Dividers-by-20/21/22/23/24	Maximum frequency = 2.4 GHz (at 125°C) Power consumption = 140 mW CMOS compatible
	Data-acquisition system	8-channel data acquisition at 1 GHz (front-end)
Telecommunications systems	4-to-1 multiplexers	Maximum bit rate = 3 Gbit/s Power consumption = 350 mW ECL compatible
	1-to-4 demultiplexers	Maximum bit rate = 3.5 Gbit/s Power consumption = 500 mW ECL compatible
Super-computers	1-kbit SRAMs (fig. 2)	Access time = 1.8 ns (at 125°C) Power consumption = 350 mW Cycle time = 2 ns
	Gate arrays	Propagation delay time = 40 ps per gate (at fan-in = fan-out = 1) Fan-in delay time = 5 ps Fan-out delay time = 6 ps Interconnection delay time = 15 ps/mm Power consumption = 7.5 mW per gate
Fast signal processing	4-bit analog-to-digital converter (fig. 3)	Maximum sampling frequency = 2 GHz Power consumption = 150 mW
	8×8 -bit modified Booth parallel multiplier	Minimum operating time = 2 ns (typically 2.5 ns) Power consumption = 350 mW 800 equivalent gates

gate), so that very complex circuits could be made with a speed equivalent to that of BFL gates. The basic normally-off logic gate used in our Laboratories consists of a normally-off transistor (with $V_T = 0.1$ V), with a resistive load. The use of this type of load is justified by the low supply voltage (only 1.5 V for source and drain) and the ease of fabrication. A follower-buffer stage can also be added to make the design more flexible and reduce the sensitivity to temperature and fan-out and interwiring capacitances. Direct-coupled FET logic (DCFL) and buffered direct-coupled FET logic (BDCFL) can easily be combined on the same chip.

In the fabrication of normally-off devices, ion-implanted active layers are used directly for load resistors and they are locally recessed to fabricate transistors with a gate length of $0.7\ \mu\text{m}$. The sheet resistance of the implanted layer is controlled to within 10%, while the local standard deviation of the threshold voltage is typically less than 30 mV over 2-inch wafers with a mean V_T -value of 0.1 V. In applications such as SRAMs and LSI (large-scale integration) digital ICs, local standard deviations of the threshold voltage as low as 15 mV are required for a functional yield as high as 50%. Pioneering work at LEP on indium doping and preannealing^{[9][10]} has made this extremely uniform material available for circuit processing. These good features result not only from a higher quality of the original material, but from better control of the ion-implantation technique — especially the annealing conditions — and also from better control of the gate recess and surface passivation. These improvements have been brought about in less than five years and it is now possible to integrate up to a thousand gates (a few thousand transistors) on a single logic chip.

The cut-off frequency of normally-off MESFETs with a gate length of $0.7\ \mu\text{m}$ is typically as high as 17-20 GHz with a transconductance of up to 180-200 mS/mm. A very short propagation delay time has been observed for normally-off logic gates in medium- or large-scale integration (MSI or LSI). Typical values of the propagation delay time t_{pd} , depending on the fan-out FO and the length L_i of the interconnection

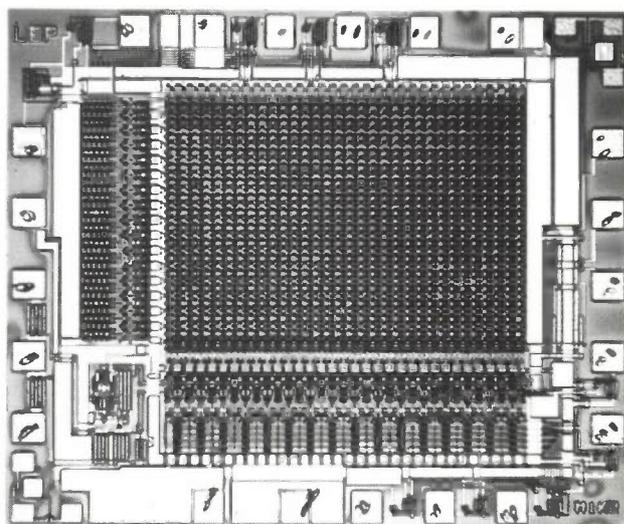


Fig. 2. A 1-kbit GaAs SRAM. The chip size is $1.9\ \text{mm} \times 2.2\ \text{mm}$. The memory-cell area is $1200\ \mu\text{m}^2$ and consists of four transistors and two resistors. The access time is as low as 4.8 ns at 125°C .

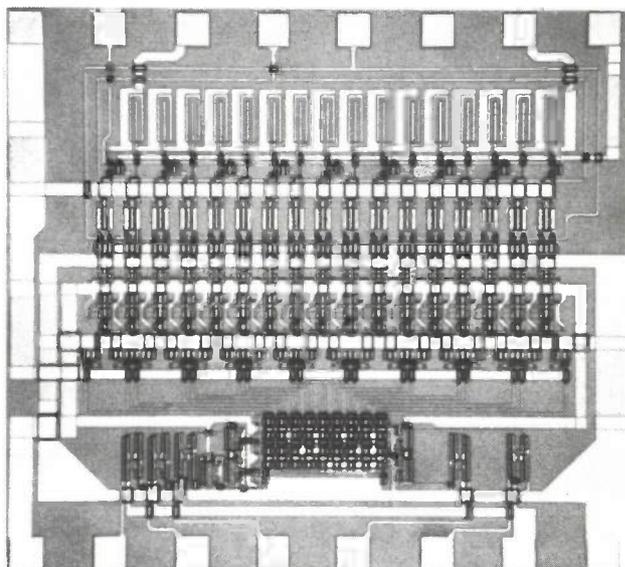


Fig. 3. A 4-bit GaAs analog-to-digital converter. This is a flash converter using fifteen latched comparators and a very compact decoder. The maximum sampling frequency is 2 GHz. The chip size is $1.4\ \text{mm} \times 1.4\ \text{mm}$.

line, are given by: $t_{pd}(\text{ps}) = 22 + 35 FO + 50 L_i$ (mm), for a DCFL gate with a power consumption of 0.4 mW per gate, and $t_{pd}(\text{ps}) = 34 + 15 FO + 20 L_i$ (mm), for a BDCFL gate with a power consumption of 2.8 mW per gate.

Applications and performance of digital GaAs ICs

The various digital applications where high-speed low-power GaAs ICs are likely to be used have now been clearly identified. *Table I* summarizes some of the applications and the performance achieved with digital GaAs ICs designed and fabricated at LEP. There are several fields of application.

[7] M. Rocchi, T. Ducourant, E. Delhaye and B. Gabillard, GaAs integrated circuits: design and technology, Collins and McMillan 1988, Chapter on digital IC design.

[8] M. Rocchi and B. Gabillard, GaAs digital dynamic IC's for applications up to 10 GHz, IEEE J. SC-18, 369-376, 1983.

[9] J. Maluenda, G. M. Martin, H. Schink and G. Packeiser, Homogeneity qualification of GaAs substrates for large scale integration applications, Appl. Phys. Lett. 48, 715-717, 1986.

[10] C. Rocher, J. Maluenda, B. Gabillard, T. Ducourant, M. Prost and M. Rocchi, Evaluation of the theoretical maximum fabrication yield of GaAs 1K bit SRAM's, Proc. 13th Int. Symp. on Gallium arsenide and related compounds, Las Vegas, Nev., 1986, Inst. Phys. Conf. Ser. 83, 509-514, 1987.

In instrumentation systems, frequency-dividers operating at high frequencies (≥ 5 GHz) are needed for counters as well as for time meters and frequency meters. More specifically, in heterodyne receivers to be used for frequency conversion, frequency synthesis is greatly simplified by using variable-modulus dividers, which can be operated up to 2 GHz at 125 °C (the highest temperature of operation specified for military application). Finally, sampling and (de)multiplexing circuits operating at 2 GHz are required for high-performance logic analysers.

Digital GaAs ICs have also high potentialities for application in telecommunications systems. They are suitable for digital links with a high bit rate

converters with a high sampling rate. Fig. 3 shows a typical example of an analog-to-digital converter^[12]. These converters are required along with sample and hold stages and, more generally, arithmetic circuits such as parallel multipliers.

Analog GaAs ICs

From hybrid to monolithic analog ICs

The evolution in time of the technologies of low-to-medium-power microwave circuits is shown schematically in fig. 4. In the sixties packaged semiconductor microwave generators appeared. There was the Gunn diode (producing high-frequency current variations

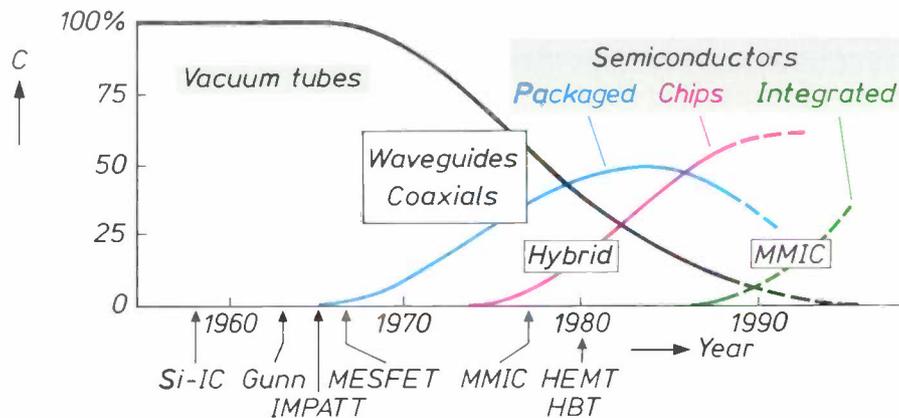


Fig. 4. Relative contribution *C* of various types of component to low-to-medium power microwave circuits through the years. The impact of thermionic valves (vacuum tubes) decreases continuously after the sixties because of the appearance of several types of semiconductor microwave devices, as indicated along the time axis. Three important transitions can be seen: the use of packaged semiconductor elements in the sixties, the development of hybrid circuits with unpackaged semiconductor chips in the seventies and more recently the appearance of monolithic microwave integrated circuits (MMICs).

(≥ 3 Gbit/s) and for digital switching systems. The main MSI/LSI functions of interest are multichannel (de)multiplexers, decision circuits, laser drivers and switching matrices. Fast GaAs SRAMs can also be useful here.

In the progressive evolution of supercomputers, digital GaAs ICs can very simply complement fast silicon-based emitter-coupled logic (ECL) circuits and CMOS ICs (CMOS stands for Complementary Metal-Oxide-Semiconductor, combining NMOS and PMOS transistors). MSI/LSI gate arrays and 1-kbit and 4-kbit SRAMs with access times ≤ 2 ns are of great interest to designers of supercomputers. As an example, fig. 2 shows a 1-kbit GaAs SRAM^[11].

Finally, digital GaAs ICs can also play an important part in fast signal processing. There is a need for low-power analog-to-digital and digital-to-analog

on application of a large constant electric field) and the avalanche diode or IMPATT diode (IMPATT stands for IMPact Avalanche and Transit Time). This was followed by the GaAs MESFET, which triggered off the first transition towards miniaturized low-to-medium-power analog circuits for high frequencies. These circuits had previously been entirely dependent on thermionic valves (vacuum tubes).

A second transition occurred in the seventies, when unpackaged chip devices were mounted directly on a low-loss dielectric substrate. Thin-film distributed lines and passive lumped elements (inductors, re-

[11] B. Gabillard, T. Ducourant, C. Rocher, M. Prost and J. Maluenda, A 200-mW GaAs 1K SRAM with 2-ns cycle time, IEEE J. SC-22, 693-698, 1987.

[12] T. Ducourant, M. Binet, J.-C. Baelde, C. Rocher and J.-M. Gibereau, 3 GHz, 150 mW, 4 bit GaAs analogue to digital converter, Proc. IEEE GaAs IC Symp., Grenelle, Fla., 1986, pp. 209-212.

sistors and capacitors) were also fabricated or mounted on the same substrate. This marked the end of the conventional microwave circuits which had used waveguides and coaxial lines.

At LEP, in-house GaAs MESFETs were used in the early work on developing a hybrid technology for microwave integrated circuits (MICs). Analog functions for frequencies from 5 to 18 GHz were successfully demonstrated, e.g. oscillators using dielectric or ferrimagnetic resonators and voltage-controlled oscillators (VCOs). Efforts were then turned to a very promising consumer application: direct broadcast satellite (DBS) television reception at 12 GHz. The outdoor front-ends for this application include a low-noise FET preamplifier, an image-rejection filter to suppress image signals at undesired frequencies, a FET mixer where the input signal is mixed with a locally generated signal for the conversion to a lower frequency, and a FET local oscillator. Typical performance figures are an overall conversion gain of 30 dB with a noise figure of 3 dB. MICs are now widely used in professional systems at frequencies up to 18 GHz. MIC distributed amplifiers for frequencies up to 30 GHz have recently been demonstrated at LEP.

A third transition is now well under way. Since GaAs substrates are semi-insulating, fully monolithic microwave integrated circuits (MMICs) can readily be fabricated. This gives new advantages such as reduction in size, suitability for large-scale production with a lower cost per unit, better matching between devices, higher circuit complexity and more accurate device modelling and performance predictions. However, MICs are still very popular because of some of their practical advantages: they are easy to trim and tune manually to meet the specifications. Their fabrication does not require high-level investment and they are well-suited to small-scale applications.

Monolithic processes for microwave circuits

Processes for fabricating MMICs like those developed at LEP are basically similar to the processes for digital ICs. Active devices are normally-on and normally-off MESFETs with recessed gates with lengths ranging from 1 μm down to 0.5 μm . The doping profile beneath the gate is optimized for low-noise applications, linear amplifiers and power devices. Additional features required for MMICs are distributed elements (microstrip lines and junctions) as well as lumped elements, e.g. metal/insulator/metal capacitors, interdigitated capacitors, high-precision metal resistors, plated via holes through the substrate for low-inductance grounding, and implanted bulk resistors.

In recent years statistical computer-aided design (CAD) models of active and passive elements have been developed for accurate time- and frequency-domain simulations, and post-fabrication trimming is now unnecessary.

Applications and performance of GaAs MMICs

Table II shows applications and performance figures for some of the analog single-function and multifunction chips which have been designed, fabricated and

Table II. Applications, functions and measured performance of some monolithic microwave GaAs ICs. The measured data agree fairly well with data obtained by computer simulation.

Applications	Functions	Performance
Television	Tuner with active image rejection (fig. 5)	Bandwidth = 100-950 MHz Conversion gain = 22 dB Image-frequency rejection = 20-40 dB
Radio link	Biphase modulator (fig. 6)	Bandwidth = 2-3 GHz Carrier rejection ≥ 40 dBc Intermodulation product ≥ 38 dBc
	Homodyne receiver	Bandwidth = 0.04-4 GHz Conversion gain = 12.5 dB Third-order intercept point = 15 dBm
Test vehicle	High-reproducibility amplifier	Bandwidth = 6-13 GHz Gain = 4.5 dB Tracking = ± 0.2 dB (circuit reproducibility)
Frequency-modulated continuous-wave radar	One-chip microwave front-end	Output power = 15 dBm Tuning linearity = 5-10% External Q = 5000
General purpose	Gyrator	Frequency = 0.9 GHz Q = 200
	Distributed amplifier (MESFET, fig. 7)	Gain = 6 ± 1 dB Bandwidth = 1-19.5 GHz
	Distributed amplifier (HEMT)	Gain = 5.5 ± 1.0 dB Bandwidth = 2-42 GHz
Electronic counter-measure systems	2-6 GHz amplifier	Bandwidth = 2-8 GHz Gain = 23.5 ± 1.5 dB Standing-wave ratio (in, out) = 1.6
Satellite television systems	Monolithic DBS front-end (12 GHz)	Bandwidth = 11.7-12.5 GHz Conversion gain = 25 ± 3 dB Noise figure = 4.5 dB
Instrumentation	Normally-off amplifier	Bandwidth = 0.1-8 GHz Gain = 8 dB Standing-wave ratio (in, out) = 1.2
Transmission/reflection modules of phased-array antenna	Power amplifier	Bandwidth = 7.7-10.1 GHz Linear gain = 28 ± 1 dB Power gain = 28.8 dB

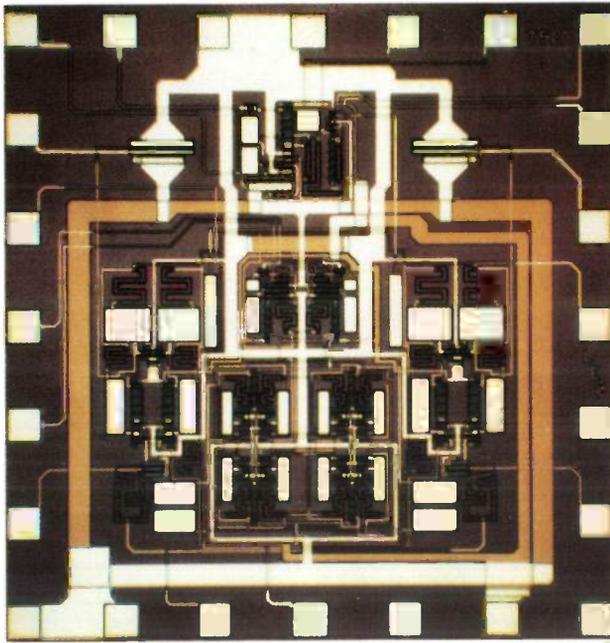


Fig. 5. A GaAs monolithic microwave integrated circuit (MMIC) for a VHF/UHF television tuner with active image rejection. The chip size is $1.5 \text{ mm} \times 1.5 \text{ mm}$.

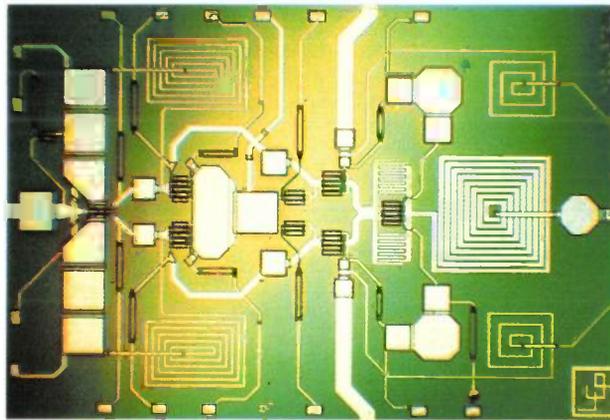
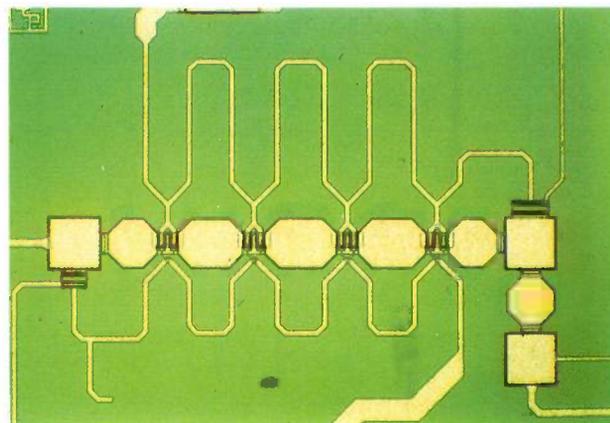


Fig. 6. A GaAs MMIC for a biphase modulator (2-3 GHz) for radio links.



tested at LEP. The type of element to be used for a given application is largely determined by the operating frequency or frequency range.

For frequencies below $f_T/3$ (f_T = unity-current-gain cut-off frequency), the wavelength is large compared with the dimensions of lumped elements, so it is not necessary to use distributed elements. Low-frequency design techniques with high packing density are used for applications such as the VHF/UHF television tuners (0.1-0.9 GHz) shown in *fig. 5*, DBS indoor units (0.9-1.3 GHz), mobile-radio systems (1 GHz), global positioning systems (2 GHz) and radio links (0.9-4 GHz). *Fig. 6* shows an example of an analog MMIC circuit to be used as a biphase modulator in radio links.

For frequencies above $f_T/3$, distributed lines are combined with lumped elements to give sufficient gain at frequencies up to f_{max} . *Fig. 7* shows a distributed amplifier for general-purpose applications (1-19.5 GHz). Besides 12-GHz DBS front-ends^[13], a consumer application, the main applications of such high-frequency devices are in military equipment such as transmitter and receiver modules in phased-array systems at 20 and 30 GHz and millimetre-wave receivers at 94 GHz.

In applications like dual phase modulators, excellent matching is possible between transistors in differential amplifiers. The MMIC version therefore gives much better performance than its hybrid counterpart currently used in these systems^[6]. Along with the possible high packing density of MMICs, this should naturally lead to new circuit architectures which cannot be made with MIC design techniques.

Heterodevices and next-generation ICs

Various high-frequency devices based on AlGaAs/GaAs or AlInAs/GaInAs heterostructures have been developed at LEP with our own epitaxial layers deposited by metal-organic vapour-phase epitaxy (MOVPE). This has resulted in the demonstration of various novel devices:

- high-electron-mobility transistors (HEMTs), in which the conduction electrons and their donor atoms are spatially separated to give a substantial increase in electron mobility at room temperature,

<

Fig. 7. A GaAs distributed amplifier with a gain of 5 dB from 1 to 19.5 GHz. The input and output distributed lines can clearly be seen. It should also be noted that the d.c. biasing networks for the gates and drains of the four transistors have been monolithically integrated. Finally the source pads are grounded via holes through the substrate.

- MIS-like hetero-FETs, where MIS stands for Metal-Insulator-Semiconductor,
- junction FETs, where a rectifying p-n junction is used for the gate electrode, and also
- heterojunction bipolar transistors (HBTs).

Table III shows applications and performance figures for heterodevices designed and fabricated at

Outlook

Since 1974, digital and analog GaAs ICs have been through a 'hard-fought war' with silicon ICs. In MICs and MMICs, the clear-cut advantage of GaAs MESFETs in terms of gain and noise figure above 1 GHz is no longer in question. Digital GaAs MESFET ICs have also been clearly shown to have a 'power-delay

Table III. Characteristic feature size l , unity-current-gain cut-off frequency f_T , maximum frequency f_{max} , minimum noise figure F_{min} (at 12 GHz), associated gain G and application of high-frequency devices, designed and fabricated at LEP.

Device	l (μm)	f_T (GHz)	f_{max} (GHz)	F_{min} (dB)	G (dB)	Application
GaAs MESFET	0.5	25		1.6	9	Low-noise digital ICs
	0.2	34				
AlGaAs/GaAs HEMT	0.5	35	125	0.85	10	Low-noise ICs
AlGaAs/GaAs MIS-like hetero FET	1	25				
GaInAs JFET	0.5	56	120			Optoelectronics
AlGaAs/GaAs HBT	4	25	20			High-power analog ICs
	(emitter)					High-speed digital ICs

LEP. Each of these transistors has some performance features better than those of MESFETs and they are all suitable for high-frequency monolithic ICs. Unfortunately, no single type has all the desired features that would enable them to replace GaAs MESFETs.

HBTs will be required for power and high-frequency SSI/MSI digital applications and interface circuits like analog-to-digital converters and digital-to-analog converters. HEMTs are ideal for low-noise MMICs, whereas n- and p-channel MIS-like hetero-FETs are best suited for CMOS-like digital circuits for operation at clock frequencies as high as 10 GHz. Heterodigital ICs and MMICs based on heterodevices are expected to reach frequencies about two to three times higher than those obtainable with conventional MESFETs at an equivalent power-consumption level. With MIS-like hetero-FETs, the fabrication of (V)LSI circuits based on highly uniform hetero-epitaxial layers becomes a promising possibility. HEMT MMICs also have much to offer: we have recently demonstrated a distributed amplifier designed with our 0.5- μm HEMT process; it has a gain of 6 ± 1 dB from 2 to 42 GHz, and the noise figure is less than 5.2 dB up to 18 GHz.

product' at least three times smaller than that of silicon equivalents with similar feature size. Further improvements in heterodevices such as HEMTs and HBTs are now expected, and these should be decisive.

We have to bear in mind, however, that the real issue is still the economic viability of submicron devices in GaAs or silicon. The investment required grows exponentially as circuit-feature size is reduced below 1 μm and it will be essential to find cost-effective ways of crossing the 0.5- μm barrier.

Summary. In GaAs MESFETs (MEtal-Semiconductor Field-Effect Transistors), the gate metal is directly applied to the GaAs surface to form a Schottky-barrier diode. The transit time of electrons beneath the gate electrode can be very short (down to a few ps) so that these transistors have very good high-frequency (> 1 GHz) characteristics. During the last fifteen years monolithic circuits based on GaAs MESFETs have been extensively investigated at LEP. Significant progress in device technology has greatly widened the field for both digital and analog applications. High-speed low-power digital GaAs ICs can be used in instrumentation and telecommunication systems, in supercomputers and in systems for fast signal processing. Analog GaAs ICs can be used in various r.f. and microwave devices, not only for professional applications at frequencies up to 30 GHz, but also for consumer applications such as the 12-GHz front-ends for direct satellite television reception. The future of GaAs MESFET ICs will depend closely on the evolution of special semiconductor heterostructures for high-frequency devices and also on the economic effects of the submicron miniaturization of GaAs devices and their silicon counterparts.

^[13] C. Kermarrec, J. Faguet, B. Vancon, C. Mayousse, A. Collet, P. Kaikati and D. Beaufort, The first GaAs fully integrated microwave receiver for DBS applications at 12 GHz, Proc. 14th Eur. Microwave Conf., Liège 1984, pp. 749-754.

Power integrated circuits

M. Amato, G. Bruning, S. Mukherjee and I. T. Wacyk

Microelectronics has made a great impact on the development of electronic systems, especially in signal processing and computing. This has been mainly due to the reduction in transistor sizes and the advanced fabrication techniques now available for very-large-scale integrated (VLSI) circuits. One of the factors common to these integrated circuits, however, is that they operate at low currents and low voltages (e.g. 5 V); they cannot conduct high currents or directly withstand high voltages. This means that VLSI circuitry cannot interface directly with power equipment or fulfill other common requirements for power handling. However, recent advances in power devices, analogous to the developments in VLSI circuits, have led to very much smaller dimensions for integratable power devices and to new gate-controlled device structures requiring very little drive power. These developments make an integrated circuit technology combining power devices and low-voltage control circuitry economically viable. The components made in this technology are known as Power Integrated Circuits or PICs. A special category of PICs, known as High-Voltage Integrated Circuits (HVICs), provides high-voltage interfacing at relatively low currents, generally between logic circuits and very-high-power discrete devices.

Introduction

Power IC technology is attractive for many reasons. Enhanced performance, increased reliability, and reduced cost are among the benefits that result when a number of discrete components are replaced by a PIC. Entirely new system approaches also become feasible, such as multiplexed power distribution for automobiles.

On the performance side PICs offer a number of desirable features. The noise level in power electronic circuits is usually very high as a result of fast high-voltage and current switching, e.g. 500 V in 20 ns at a 2- μ s repetition interval. In discrete circuits wiring inductances and stray capacitances tend to be sensitive to this noise and require very special care in layout and design. In a PIC, interconnecting leads are so short that the inductive pick-up is significantly re-

duced. Capacitive coupling is also reduced since the parasitic capacitances are lower.

On-chip protection such as thermal shutdown can be directly linked to the power device instead of attaching a temperature sensor to a heat sink on which a discrete power device is mounted. This is very important as this possibility demonstrates both a functional advantage and a saving in discrete components. Control circuits for over-current and over-voltage protection may now be added where economic aspects formerly prohibited this and called for overdesign. The incorporation of intelligence into power circuits can have such distinct effects that it is sometimes referred to as 'smart power' [1].

For all these reasons PICs are likely to enhance performance and have an impact on manufacturing cost by component reduction. Although the advantages are manifold there are challenges to meet as well. Owing to the increased process complexity, the cost of

M. Amato, G. Bruning, Dr S. Mukherjee and Dr I. T. Wacyk are with Philips Laboratories, North American Philips Corporation, Briarcliff, New York, U.S.A.

manufacturing PICs per unit area of silicon is higher than for discrete power devices. The cost advantage of a PIC over a discrete implementation depends on the overall system economics, where the benefit is derived from a lower package count, smaller printed-circuit board, less weight, etc. Furthermore, since the applications in power electronics are very diverse, a high degree of flexibility in power devices is required. Both difficulties are addressed by offering more control capability, smaller power device structures, increased reliability, and a fast turn-around design cycle resulting from advanced design methods.

Power electronics can be found today in any machine or device which requires a form of electricity other than that supplied by the primary source (e.g. from the utility company) and where a high degree of regulation of output voltage, output current or output power is desired. Fig. 1 illustrates the diversity of

chines, compressor motors, disk drives and shavers;

- electronic ballasts for discharge lamps, which give lower energy consumption for a given amount of light;
- automotive switches, which replace mechanical relays and expensive heavy wiring harnesses.

The interface between large discrete power devices and the common low-current, low-voltage ICs can be made by High-Voltage Integrated Circuits (HVICs), which operate at relatively low currents and high voltages. These are also indicated in fig. 1.

PIC technology is the result of a number of recent advances in microelectronics. These include integratable power devices such as the Lateral Double-diffused MOS (LDMOS) transistor and the Lateral Insulated-Gate Bipolar Transistor (LIGBT), and the development of a fabrication process which combines both low-voltage CMOS control and these power devices on the same chip. The problem of making full use of these advances has also been addressed at the same time. The construction of a structured design method permits the non-expert designer of ICs and power devices to design application-specific PICs based on a standard-cell approach and a highly automated layout technique. Examples of power integrated circuits [2] are depicted in fig. 2. The figure shows two versions of a PIC designed for a television switched-mode power supply. These circuits combine a 500-V/5-A switch with all the control and protection functions, including on-chip over-temperature protection, which are required by the television power supply. In fig. 2a an LDMOS transistor, which is the dominant structure on the chip, has been used for the power switch, whereas in fig. 2b an LIGBT is shown. The LIGBT is a relatively recent development and, because it can handle a higher current density than the LDMOS device, it clearly results in a smaller and cheaper chip for a given function. The control circuit is located under the power devices and is the same in both cases.

In the following sections we shall first give a detailed description of the devices and the technology which can be used in the fabrication of economically attractive PICs. We shall make a distinction here between high-voltage circuits and high-current circuits, which require different approaches. Then we shall explain our design methodology, since the ease of design is a decisive factor in the economic viability of PICs, and even more so for Application-Specific PICs (ASPICs).

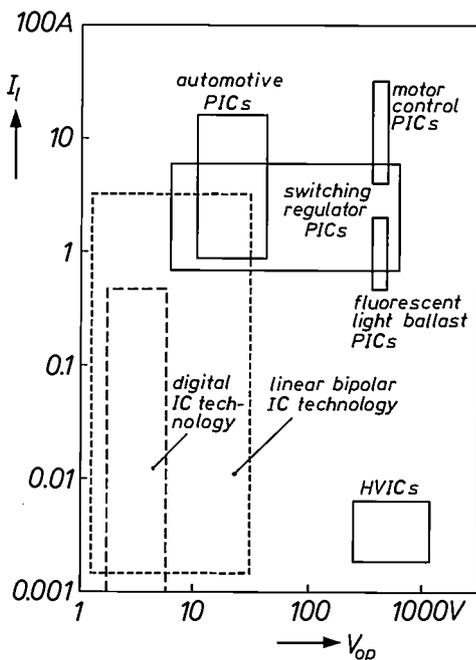


Fig. 1. Typical applications for Power Integrated Circuits (PICs) as a function of operating voltage V_{op} and load current I_i . To emphasize the differences from other IC technologies, dashed lines indicate the practical working ranges for common digital IC technology and linear bipolar IC technology. HVICs are special types of PICs which can function as interfaces between ICs fabricated in different technologies.

applications and operating requirements for PICs. Some typical examples are:

- switched-mode power supplies or switching regulators, which efficiently convert a.c. power into d.c. power and are found in almost all computers, television sets and electronic equipment;
- motor controls, which provide precise speed and torque control in applications including washing ma-

[1] D. Paxman, Smart power, Phys. World, No. 2 (January), 26-28, 1989.
 [2] S. Mukherjee, M. Amato, I. Wacyk and V. Rummennik, LDMOS and LIGBT's in CMOS technology for power integrated circuits, Int. Electron Devices Meeting, Washington D.C., 1987, Tech. Digest, pp. 778-781.

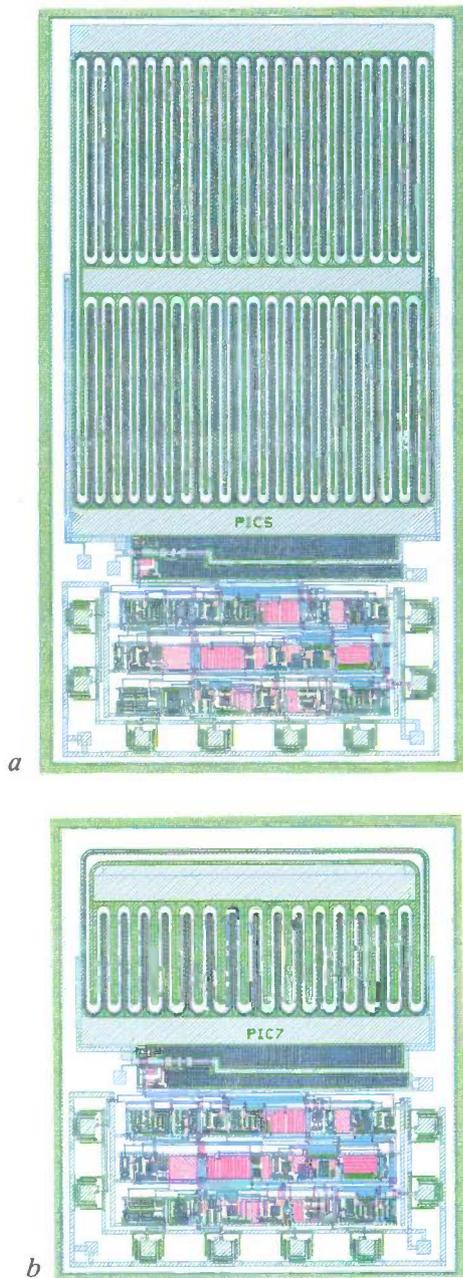


Fig. 2. Two examples of a PIC for application in a television switched-mode power supply. The large structures are the power transistors. The intricate, finely detailed structures are the signal-processing circuitry for control and protection and are the same for both ICs. *a)* The power transistor here is of the Lateral Double-diffused MOS (LDMOS) type; the area is 22 mm². *b)* When a Lateral Insulated-Gate Bipolar Transistor (LIGBT) is used, the total area can be reduced to 11 mm².

Devices and technology

Research on semiconductor power devices that can be used in PICs is concentrated on two main device groups:

- high-voltage devices (breakdown voltage > 250 V), and
- high-current devices (on-current > 5 A).

In both, reducing the size of the power device for the same function is a main objective of the research. In many applications power devices occupy more than half the area of the PIC chip. This means that reduction of power-device size is a key to increased performance per unit cost.

High voltage

The high-voltage PIC technology developed at the Philips Research Laboratories in Briarcliff (New York, U.S.A.) combines low-voltage CMOS control circuits with high-voltage transistors on a single silicon chip. High-voltage devices can be made as vertical or lateral structures in PIC technology. Integration of vertical high-voltage transistors and low-voltage control circuitry requires the use of thick or multiple epitaxial layers, or complicated and expensive dielectric insulation. Lateral power-device integration, on the other hand, requires only a thin epitaxial layer (5-10 μm), which is more compatible with standard low-voltage process technology.

Normally the high-voltage transistors are operated as switches. For low power dissipation, the high-voltage transistors must be capable of blocking high voltages with small leakage currents when 'off' and have a low forward-voltage drop when the switch is 'on'. Two types of lateral high-voltage output transistors are used: the LDMOS and the LIGBT. These transistors are capable of operating over a wide range of voltages (up to more than 1000 V), frequencies (up to several megahertz) and currents (up to several amperes).

The integrated LDMOS transistor is a MOS-based device with a lateral buffer zone (drift region) between the active channel and the output (drain) of the transistor. When the transistor is 'off' the drift region depletes, so that the voltage decreases laterally across the transistor. A cross-section of an LDMOS transistor is shown in *fig. 3*, indicating the high-voltage drift region associated with the device. The depletion of the drift region is a two-dimensional effect, which is called RESURF (for REDUCED SURFACE electric fields^{[3][4]}). This effect is controlled during the manufacture of the IC by ion implantation and diffusion of the moderately doped drift region. LDMOS power transistors, unlike bipolar transistors, are majority-carrier devices and are capable of operating at very high frequencies. However, the drift region acts as a high resistance when the device is 'on', and higher-voltage devices require longer drift regions. The drift geometry, drift doping, buried layer, and substrate doping levels must be optimized to give the highest breakdown voltage with the lowest resistance for a given silicon area.

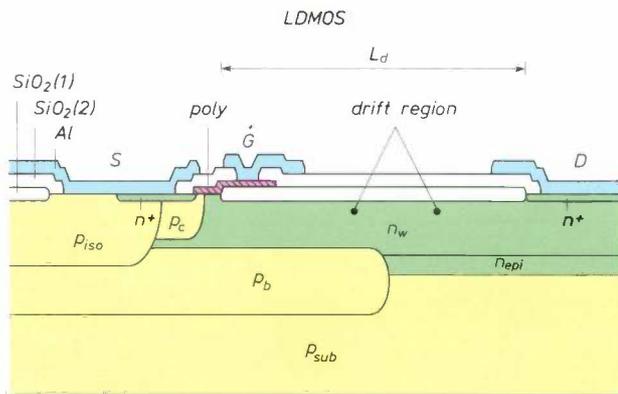


Fig. 3. Cross-sectional view of a Lateral Double-diffused MOS (LDMOS) device. *S* source, *G* gate, *D* drain. The drift region has a length L_d . The significance of the other designations is:

- p_{sub} p⁻-type substrate
- p_b p-type buried layer
- p_c p-type diffusion for channel
- p_{iso} p⁺-type isolation diffusion
- n_{epi} n-type epitaxial layer
- n_w n-type well
- n^+ n⁺-type diffusion for source and drain
- Al* aluminium metallization
- Poly* polysilicon layer
- SiO₂(1)* silicon 'field' oxide from LOCOS processing
- SiO₂(2)* silicon isolation oxide

(Between the conducting polysilicon layer and the semiconducting channel diffusion there is always a very thin insulating SiO₂ layer of some tens of nanometers; this is not shown explicitly here or in the following figures.)

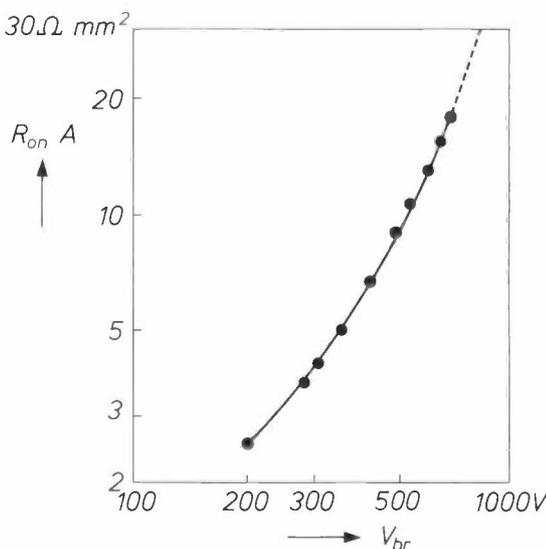


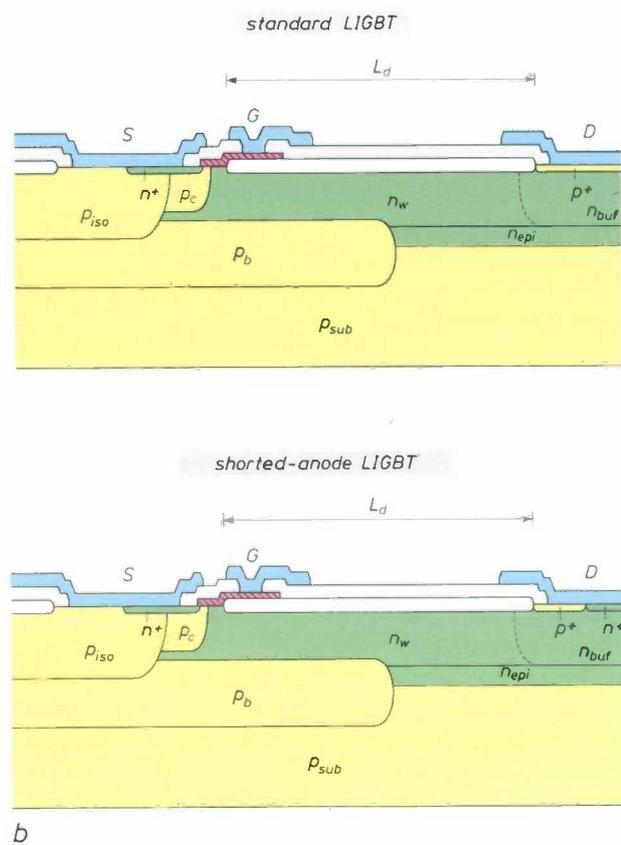
Fig. 4. Specific on-resistance $R_{on} \cdot A$ plotted against breakdown voltage V_{br} for a typical LDMOS transistor (simulated results); R_{on} represents the resistance in Ω and A the silicon area of the transistor in mm^2 .

The specific on-resistance of a transistor is defined as the resistance at a specific breakdown voltage for unit area of silicon. A lower specific on-resistance (indicated by $R_{on} \cdot A$ and with the dimension $\Omega \cdot \text{mm}^2$) means a more efficient design which translates into a smaller chip size for specified power dissipation and a lower silicon cost. The relationship between specific on-resistance and breakdown voltage for a typical integrated LDMOS transistor is shown in *fig. 4*. When a high current and a low power dissipation are required, the LDMOS transistors are laid out with the source and drain regions interdigitated, increasing the channel width and reducing the total on-resistance. This can be seen in the diagrams of the switched-mode power supply chip in *fig. 2*. The LDMOS transistor is best suited for high-frequency applications where the r.m.s. load current is less than 1 A.

Although the LDMOS has been one of the pioneering MOS power devices in PICs for high-voltage applications, the major limitation of this device is a relatively high specific on-resistance due to the majority-carrier conduction. For high voltages ($> 250 \text{ V}$) the major contributor to the on-resistance is the drift region shown in *fig. 3*. Since the drift region is designed to sustain the drain voltage in the off-state without reaching the critical breakdown field in silicon, the doping of this region is limited by the RESURF conditions. In order to overcome this limitation the LIGBT was developed^[5]. In the LIGBT, a minority-carrier injector is provided at the drain end as shown in *fig. 5a*. The injection of minority carriers into the drift region in the on-state modulates the conductivity of the drift region and thereby reduces the effective on-resistance of the transistor. Therefore, for a similar device size, the LIGBT exhibits a lower on-resistance than the LDMOS by a factor of 5 or more depending upon the breakdown voltage and current level. This device can also be viewed as a MOS-gated bipolar transistor. In the LIGBT, the minority-carrier injection, or the bipolar transistor turn-on, occurs above a certain drain voltage (0.6 V) required for forward bias of the injector junction. Below this voltage the device does not turn on, producing an additional nonlinearity in the current/voltage characteristics as compared with the LDMOS shown in *fig. 6*. However, in switching applications this nonlinearity does not pose a limitation except for setting a lower limit to the voltage drop across the device in the on-state.

The injection of minority carriers in the LIGBT device results in increased turn-off delay (1-3 μs) for the structure shown in *fig. 5a* as compared with the LDMOS (10 ns), because of storage-time effects. An improvement in the turn-off time of the LIGBT is

[3] J. A. Appels and H. M. J. Vaes, High voltage thin layer devices (RESURF devices), Int. Electron Devices Meeting, Washington D.C., 1979, Tech. Digest, pp. 238-241.
 [4] S. Colak, Effects of drift region parameters on the static properties of power LDMOST, IEEE Trans. ED-28, 1455-1466, 1981.
 [5] M. Darwish and K. Board, Lateral resurfed COMFET, Electron. Lett. 20, 519-520, 1984.



b

Fig. 5. *a*) Cross-sectional view of a Lateral Insulated-Gate Bipolar Transistor. The device looks very much like an LDMOS (see fig. 3), but the drain diffusion is now of the p^+ -type and is made in a region called n_{buf} which is intended to prevent punch-through from the drain to the substrate. The combination of drain and drift region functions as an injector junction. Conduction in the on-state is therefore due to both majority carriers and minority carriers, which reduces the on-resistance by a factor of 5 or more. *b*) Introducing an additional shorting n^+ -diffusion at the drain gives a 'shorted-anode LIGBT' or 'clamped-anode LIGBT'. This device combines the advantages of the LDMOS transistor and the LIGBT.

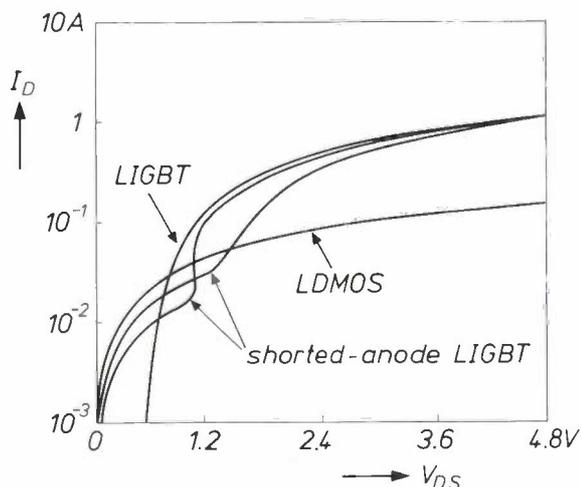


Fig. 6. Typical characteristics showing drain current I_D plotted against drain/source voltage V_{DS} for an LDMOS transistor, an LIGBT and two shorted-anode LIGBTs.

achieved by incorporating a shorting n^+ -diffusion at the drain as shown in fig. 5b^{[6][7]}. The addition of this shorting diffusion results in more efficient removal of electrons thereby decreasing the turn-off time to less than 330 ns. Additionally, the shorting n^+ -diffusion enables conduction of this 'shorted-anode LIGBT' prior to the turn-on of the injector junction in a manner similar to the LDMOS, thereby reducing the voltage drop across the device and the power loss during switching. The d.c. current/voltage characteristics of the shorted-anode LIGBT are shown in fig. 6 for two cases, together with those of the conventional LIGBT and LDMOS. From this figure it is clear that at high forward voltages the current level of the shorted-anode LIGBT is similar to that of the conventional structure because of the dominance of the conductivity modulation and bipolar conduction.

The high-voltage 'Briarcliff' IC process is based on a junction-isolated technology utilizing a relatively thin (6-7 μm), high-resistivity, n-type epitaxial layer. The process architecture is based on a dual-well low-voltage CMOS process with self-aligned n-wells and p-wells, as shown in the cross-section of fig. 7. The p^- -substrate ensures good high-voltage operation in the lateral LDMOS and LIGBT devices, while the buried n^+ -layer provides latch-up suppression in the CMOS control portions of the chip. The n-well is used for the channel in the low-voltage PMOS transistors and for the drift region in the LDMOS and LIGBT devices. The buried p-layer provides extra field shaping in the LDMOS and LIGBT high-voltage devices. A heavily doped p^+ -diffusion in source areas of the high-voltage transistors provides latch-up suppression in the LIGBT, better protection from inductive surges and insulation from the low-voltage circuitry. However, it is not necessary to provide this insulation in the control-circuit sections of the chip. The technology features self-aligned-to-polysilicon NMOS, PMOS and LDMOS/LIGBT channels on 600 \AA (60 nm) of gate oxide. It also features standard LOCOS processing, 3- μm design rules, and two polysilicon ('poly') layers for high-quality poly/poly capacitors and insulated poly resistors.

Several issues need to be considered in combining LIGBTs with CMOS devices. The interaction of minority carriers generated by the LIGBT with the CMOS low-voltage sections is one of the most important considerations. In this respect the shorted-anode LIGBT described earlier, turns out to offer a particularly promising approach. The anode short, combined with the buried layer and the deep isolation diffusions, helps to confine the minority carriers close to the power device and extract them efficiently. This, combined with n-type and p-type buried layers in the

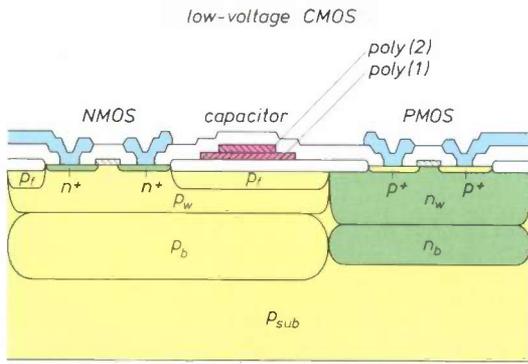


Fig. 7. Cross-section of low-voltage CMOS devices that can easily be combined with the high-voltage devices of fig. 3 and fig. 5. This figure shows an NMOS transistor, a PMOS transistor and an integrated capacitor. The significance of the designations not given earlier is: P_w p-type well, P_f p-type 'field' diffusion, n_b n⁺-type buried layer. The field diffusion is also called a 'channel-stop' diffusion because it should prevent the formation of parasitic channels.

channels on the silicon surface, thereby using up silicon area, the silicon trench provides the possibility of constructing the MOS transistor channel along the trench side wall yielding a Trench Double-diffused MOS (TDMOS) transistor (fig. 8b) [8]. This in effect provides a pseudo-three-dimensional integration, thereby increasing the silicon area efficiency (more channel width for the same silicon area). Additionally, by providing an even more pronouncedly vertical current path the pinch resistance is dramatically reduced as compared with conventional VDMOS transistors. The combination of these factors results in a substantial reduction in specific on-resistance.

The above approach has been used in a five-mask process for fabricating TDMOS transistors with rectan-

low-voltage sections, has been found to be adequate in eliminating the interaction problems with low-voltage CMOS circuits.

High current

In high-current applications the Vertical DMOS (VDMOS) is the most commonly used MOS gate-controlled power device. In this type of device the drain is situated under the source (fig. 8a), hence the term 'vertical'. The major contributors to the on-resistance in this device are the channel resistance and the pinch resistance. The reduction of these two components is therefore a primary goal and is most conveniently achieved by reducing the cell size by advanced lithography and minimizing the spacing between the adjacent channel regions also known as the pinch or JFET region (JFET = Junction Field Effect Transistor). Reducing the size of the pinch region helps to reduce the size of the cell but also tends to increase the pinch resistance. Therefore an optimization is called for.

With the advent of silicon trench-etching techniques using reactive ion etching (RIE) an alternate approach becomes feasible. Instead of fabricating the

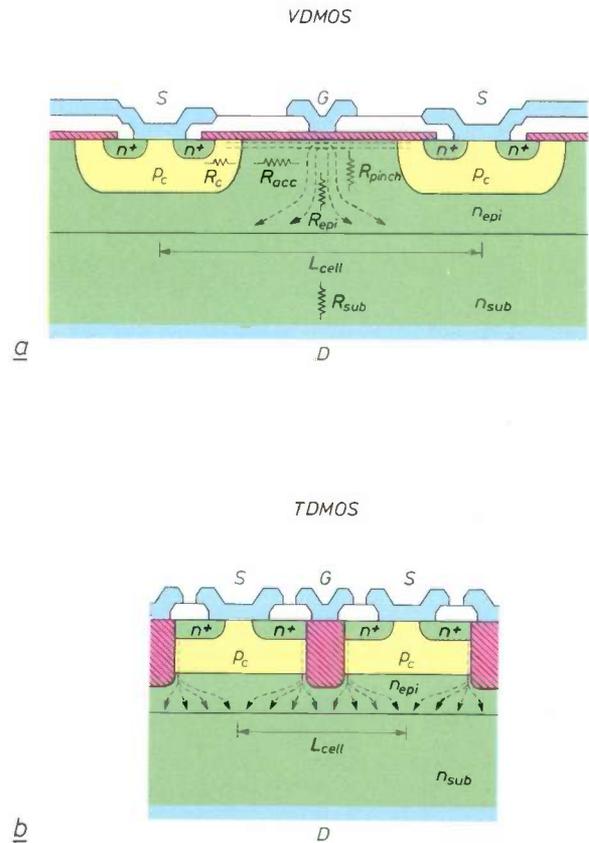


Fig. 8. a) Cross-section of a Vertical Double-diffused MOS (VDMOS) device. A part of a periodic (e.g. interdigitated) structure is shown here, so that multiple sources can be identified. The basic unit is called a cell and the cell pitch is indicated by L_{cell} . The drain is situated beneath the sources and the electron flow in the conducting state is represented by dashed arrows. A number of resistances affect the magnitude of the electron flow: the channel resistance R_c , the accumulation resistance R_{acc} , the pinch resistance R_{pinch} , the epitaxial-layer resistance R_{epi} and the resistance R_{sub} of the n⁺-type substrate n_{sub} . **b)** An increased channel density per unit area of silicon and considerable reduction of the pinch resistance can be obtained by using trench-etching techniques. Here a cross-section of the resulting Trench Double-diffused MOS (TDMOS) transistor is shown. The channel is no longer situated on the surface of the silicon crystal, but vertically, next to the gates.

[6] R. Jayaraman, V. Rumennik, B. Singer and E. H. Stupp, Comparison of high voltage devices for power integrated circuits, Int. Electron Devices Meeting, San Francisco, Cal., 1984, Tech. Digest, pp. 258-261.
 [7] P. A. Gough, M. R. Simpson and V. Rumennik, Fast switching lateral insulated gate transistor, Int. Electron Devices Meeting, Los Angeles, Cal., 1986, Tech. Digest, pp. 218-221.
 [8] S. Mukherjee, M. Kim, L. Tsou and M. Simpson, TDMOS — an ultra-low on-resistance power transistor, IEEE Trans. ED-35, 2459, 1988.

gular cell geometry. A Scanning Electron Microscopic (SEM) cross-section of one of the transistors is shown in *fig. 9a*. *Fig. 9b* shows the current/voltage characteristics of this transistor. The specific on-resistance for the device is $50 \text{ m}\Omega \cdot \text{mm}^2$ with a breakdown voltage of more than 60 V. For comparison, the specific on-resistance that can be achieved with conventional VDMOS for the same voltage range is several times higher than this.

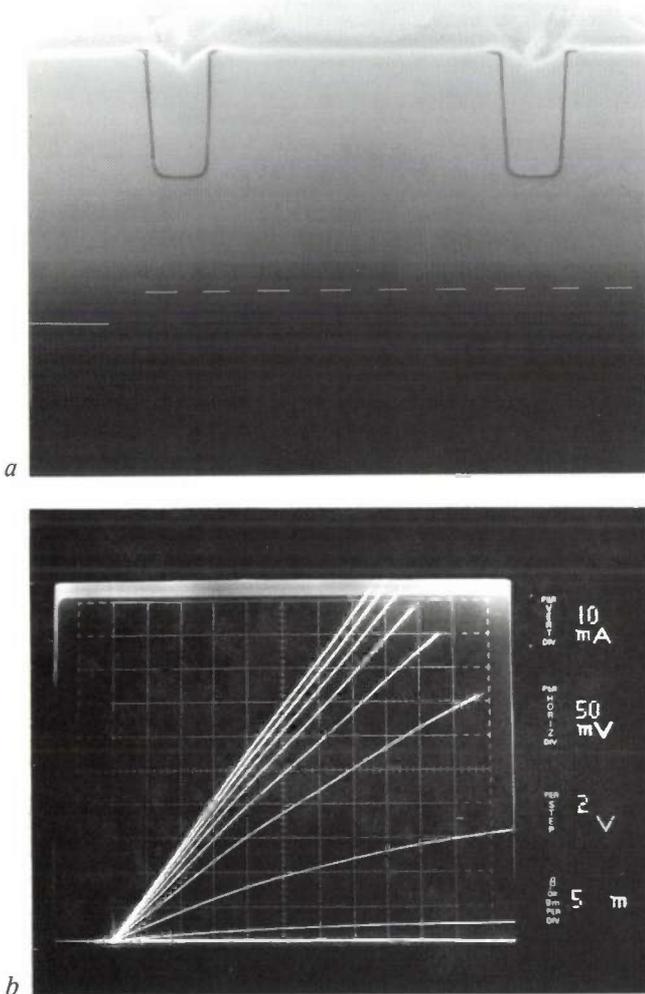


Fig. 9. *a*) Cross-section of a TDMOS transistor made by scanning electron microscopy (SEM); 1 scale division = $1 \mu\text{m}$. *b*) Measured characteristics showing drain current plotted against drain/source voltage for this device; 1 vertical scale division = 10 mA; 1 horizontal scale division = 50 mV; the gate/source voltage varies in steps of 2 V.

In seeking further improvement of TDMOS devices, several factors have to be borne in mind. First, the current conduction along the trench side wall results in carrier (electron) mobility degradation and consequent increase in channel resistance. This degradation is strongly dependent upon the silicon crystal lattice orientation, trench-etching technology and the pre-gate oxidation cleaning processes. Among

the other parameters of importance, one of the most significant parameters is the fixed oxide charge on the trench side wall. This is also a function of the pre-oxidation cleaning process as well as the oxidation conditions.

The drain/source breakdown voltage of the TDMOS transistor is determined by a combination of n-epi and p-body doping concentration together with the trench-etching profile. Two breakdown mechanisms can be identified: planar junction breakdown and avalanche breakdown under the gate. In specific situations as in the devices described above, the breakdown voltage is determined by a combination of the above mechanisms, resulting in 'field crowding' at the trench corners. It is expected that rounded trench corners will improve the breakdown voltage in such cases. Other ways of improving the breakdown voltage involve modification of the n-epi concentration and thickness and p-body doping profile. Simulations show that by proper choice of these parameters breakdown voltage can be increased with minimal change of on-resistance.

With the maturity of the RIE trench-etching technology used in VLSI memory products, the combination of TDMOS devices with standard CMOS control appears to be a practical approach for production. The technology is also compatible with standard VDMOS devices.

Design methodology

A major challenge to the widespread application of PICs is the difficulty of translating a discrete power-system design into a functionally equivalent IC. A power system usually interfaces to a complex electrical load and is required to deal with real-world analog voltages and currents. Because it is not easy to model and simulate complex systems of this type, the system engineer traditionally addresses the problem by building a breadboard prototype, and arrives at a solution by a process of testing and redesign. However, the resulting design may not be well suited for integration, and may require a considerable development effort to achieve an effective IC implementation. In addition, while logic ICs are designed for standardized electrical operating parameters, the operating voltages and currents for a PIC depend on the specific application and can range over several hundred volts and from a fraction of an ampere to tens of amperes. So although there will be a number of standard-product PICs developed for large-volume applications, a great number of the requirements will be specific to a particular application requiring a customized IC implementation.

To address the requirements of designing PICs, a methodology has been evolved in which advanced CAD techniques are used in the design of the Application-Specific Power IC (ASPIC) [9]. The objective is to make PIC design available to the engineer who is not necessarily an expert IC designer, so that power systems are developed and optimized for integration right from the start of the design process. A second objective is to make PICs economical for the large number of applications that only require a moderate volume of products. This is made possible by employing advanced CAD techniques to reduce the time and engineering effort required to produce customized silicon chips. The design methodology comprises three major facets:

- structured design techniques;
- a knowledge base of PIC design;
- customized software tools for computer-aided design and computer-aided engineering (CAD/CAE).

First, the overall methodology is based on a framework of structured design techniques that are similar to those successfully employed in digital VLSI. These techniques have been extended to address the mixed analog/digital/power environment for circuits used in PIC designs. As in any complex system, the key to successfully managing the complexity is to partition the system into smaller modules or building blocks, so that the design can take place at a high level of abstraction. The specific low-level details of device and circuit implementation are treated by experts in these areas. In addition, the building blocks have been developed in accordance with a set of rules that govern their interface properties and lead to predictable behaviour when the modules are combined into larger networks. These rules, together with the computer programs that execute them, form the basis of the structured design. Several levels of abstraction are used in the PIC design process illustrated in *fig. 10*. At the top of the hierarchy is the system diagram which consists of the PIC represented as a functional block, circuit models for the external sources and loads, and other external components. The PIC itself can be partitioned into various levels of increasing detail until the transistor-level circuit is finally reached. The rules of the structured design are carried down to the level of the primitive building blocks, or standard cells, which embody both analog and digital operations, and are described by simple functional relationships. An application is customized by structuring the known functions to solve the system problem. A benefit of designing a system at a level of abstraction re-

moved from the detailed circuit and device structure is that when the cells are upgraded as a result of technology scaling, the system can easily be 'recast' in the new process, e.g. for improved economy or performance.

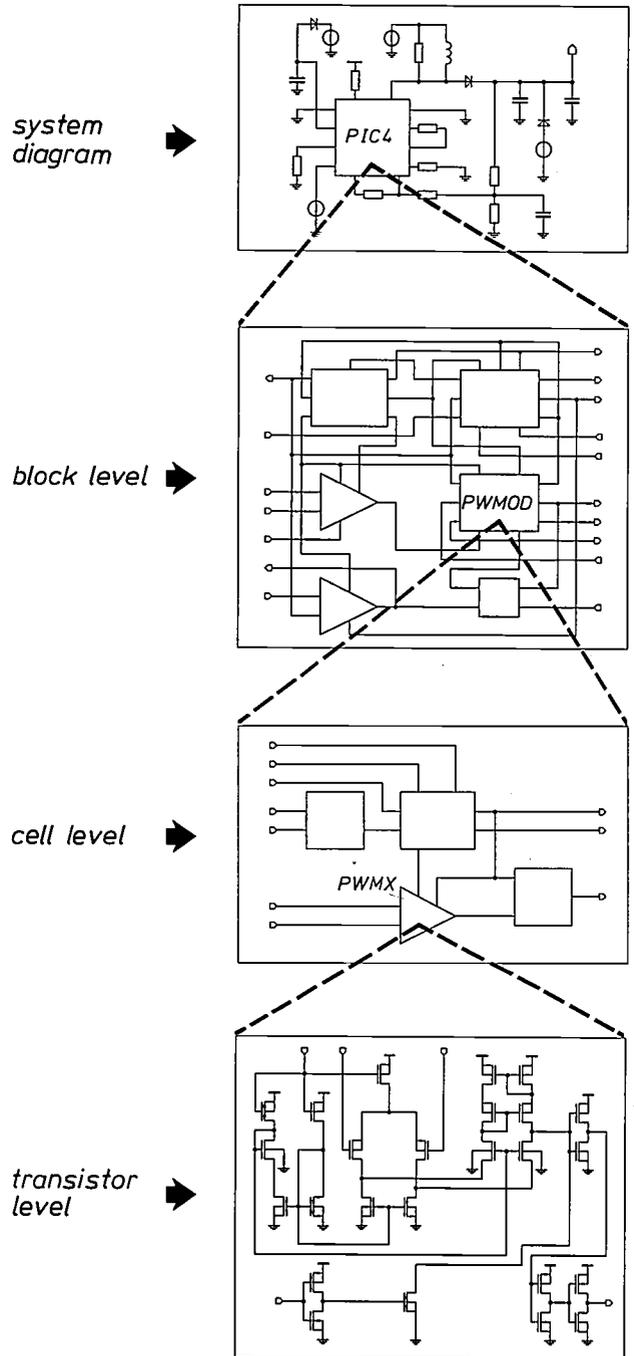


Fig. 10. PIC structured design hierarchy. Several levels of abstraction are shown, with varying degrees of detail. The design at each level can be carried out independently, as long as certain explicit design rules are obeyed. The highest level is the diagram of the system in which the PIC is to be applied. Then we have the block diagram of the PIC. One step lower we have the standard cells from which the blocks of the PIC are built up and the lowest level consists of the electrical circuits of the standard cells.

[9] I. Wacyk, M. Amato and V. Rumennick, A power IC with CMOS analog control, IEEE Int. Solid-State Circuits Conf., Anaheim, Cal., 1986, Digest of Tech. Papers, pp. 16-17.

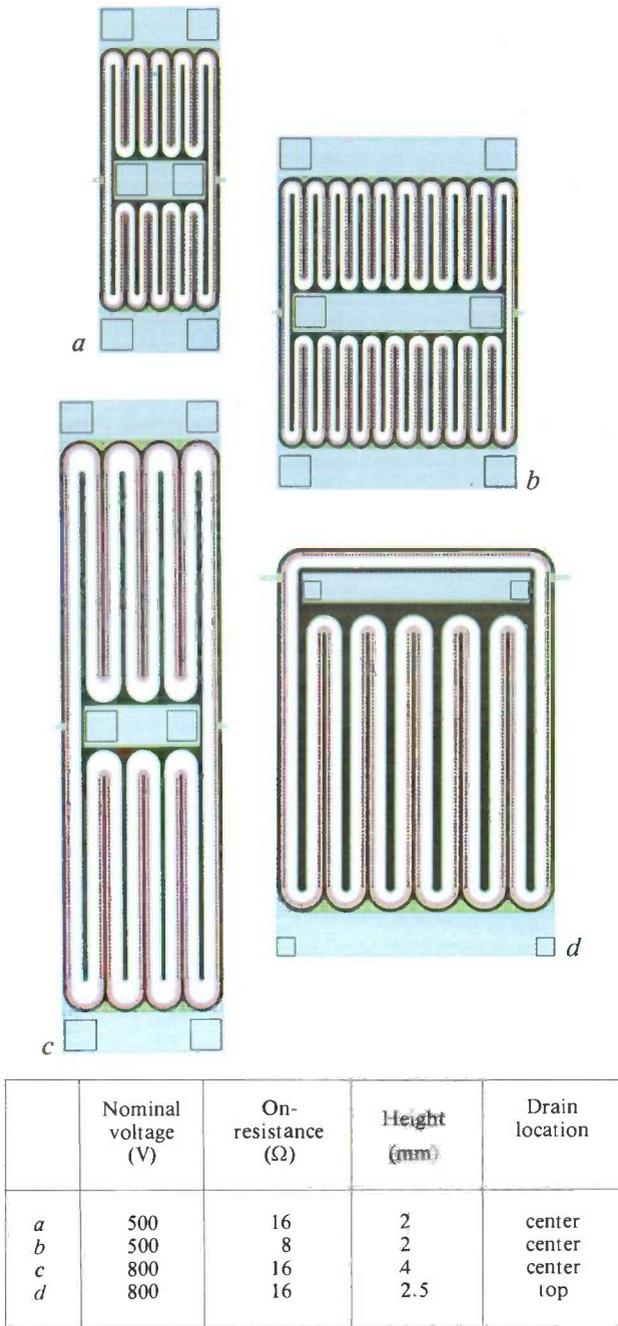


Fig. 11. Examples of several high-voltage transistor layouts, obtained ('compiled') from different sets of input parameters with the design program PROCEDURAL HIGH-VOLTAGE DEVICE LAYOUT GENERATOR. This program produces a complete sequence of layout mask files for each transistor in a few minutes. The main input parameters for the four cases are shown in the table.

Secondly, a knowledge base of PIC design in terms of power devices and power-related circuit functions is captured in computer programs and in the computer database. These form a catalog of building blocks that contain detailed device layout structure and specific circuit implementations of analog and digital functions in the target technology. One example of such a program is the PROCEDURAL HIGH-VOLTAGE

DEVICE LAYOUT GENERATOR [10]. Given a set of input parameters that may include the desired breakdown voltage, on-resistance, physical aspect ratio, etc., the complete mask-ready layout is generated ('compiled') in a matter of minutes. Fig. 11 shows several completed devices.

The circuit functions that make up the standard-cell library are another form of stored knowledge, and range in complexity from simple comparators and amplifiers to high-performance pulse-width modulators. They include many specialized functions for power control and protection as well as standard analog and digital types. These cells are represented in several forms in the computer database for use in different stages of the design process, as shown in fig. 12. The symbol designates the cell's functional behaviour, the circuit diagram its structural form, and the layout corresponds to its representation in the physical domain. The cells have been specifically em-

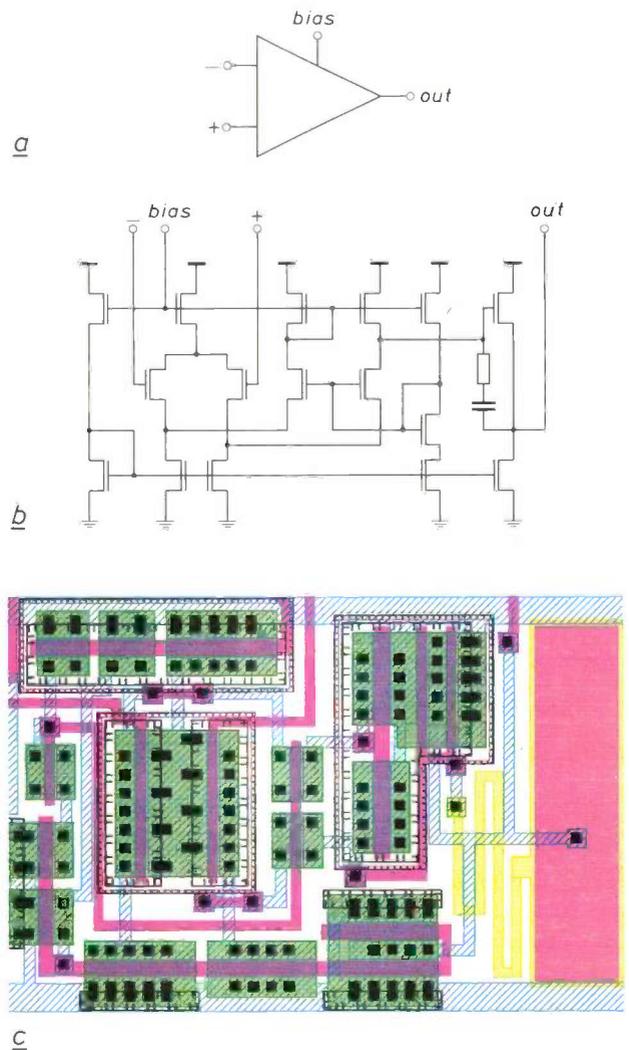


Fig. 12. The standard cells of the structured design are represented in the computer knowledge base in several forms, as shown here for a cascode operational amplifier: a) symbol, b) circuit diagram, c) layout.

bodied in CMOS technology to conform to the structured design, and to achieve low-power operation and high packing density.

Finally, specific CAD/CAE tools have been developed and customized to automate the tedious and error-prone aspects of the design process.

Fig. 13 shows the major steps in the overall design procedure. An advanced design workstation serves as

routing program is used to create a control block directly from the circuit netlist. This and the other blocks at the top of the design hierarchy are assembled and interconnected into a finished chip design with a program developed at Briarcliff called ASPICGEN. Coded in a procedural layout language, ASPICGEN invokes layout generators for various modules and automatically places and routes the high-

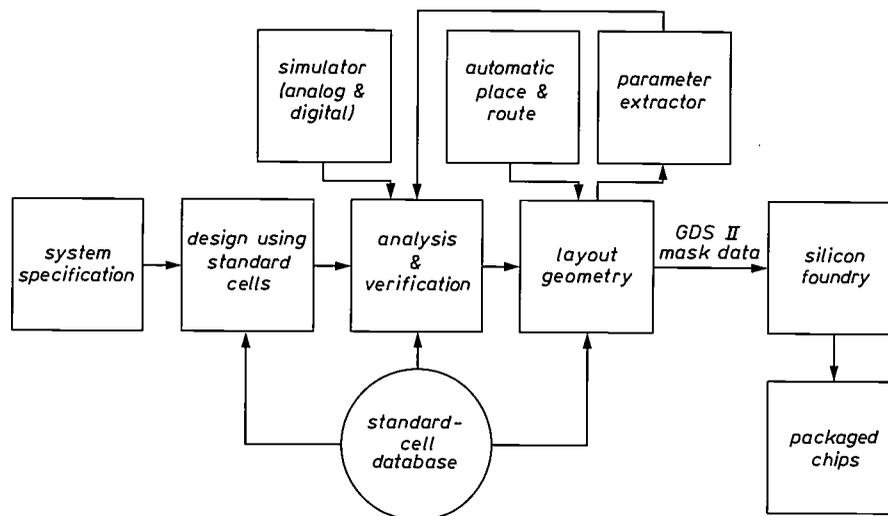


Fig. 13. Overview of the overall PIC design procedure. Most terms used in this diagram are self-explanatory. GDSII is a *de facto* standard within the IC industry for the format of layout data on magnetic tape; it refers to the format developed for and used by CALMA layout systems.

a platform for the highly integrated set of software programs. The system design is carried out using the symbolic representation of the functional blocks. A fundamental requirement for the ASPIC design methodology is the ability to verify the system operation before silicon is fabricated. A circuit-simulation program of the type known as SPICE^[11] is used to evaluate small portions of the design in detail. A second approach using the Continuous Logic (CLOGIC) simulator^[12] allows a complete system-level simulation to be carried out several orders of magnitude faster than a SPICE simulation. CLOGIC is a functional-level simulator that was developed at Philips Research Laboratories at Briarcliff specifically to address the need for verifying large-scale analog/digital circuits. In both simulators the appropriate netlists, which represent a complete survey of the interconnections of all the components, are generated automatically from the high-level system description, using the circuit representations of the standard cells. After a successful verification, the design is reduced to a physical chip layout in several steps. An automatic placement and

voltage devices, blocks of control circuitry, device drivers, and I/O pads according to a specified chip floorplan.

Two examples of a PIC design carried out with the methods described above have already been shown in fig. 2. In both cases the power device layouts were created by the compilation program described earlier, the control section was implemented using standard cells, and the final chip was assembled automatically. Intended for high-frequency switching control applications, the PIC contains a high-speed pulse-width modulator for duty-cycle control, a high-gain error amplifier for gain and compensation, voltage and

^[10] N. J. Elias, A case-study in silicon compilation software engineering, HVDEV high voltage device layout generator, Proc. 24th ACM/IEEE Design Automation Conf., Miami Beach, Fla., 1987, pp. 82-88.

^[11] L. W. Nagel, SPICE2: a computer program to simulate semiconductor circuits, Electronics Research Laboratory, Memorandum No. ERL-M520, College of Engineering, Univ. of California, Berkeley, Cal., May 9, 1975.

^[12] D. Giannopoulos, J. C. Lin, I. T. Wacyk and J. L. Woo, Circuit simulation of power ICs, IEEE Int. Solid-State Circuits Conf., San Francisco, Cal., 1988, Digest of Tech. Papers, pp. 262-263.

bias-current reference sources, and a protection and start-up sequence controller. In a switched-mode power supply operating at a switching frequency of 0.5 MHz, this chip delivered a regulated power level of nearly 100 W. The LDMOS version has been operated at a switching rate of up to 3 MHz, or more while the LIGBT version provides comparable power control in a substantially reduced chip size for frequencies of up to 100 kHz or more. Implementation of the analog and digital functions in CMOS resulted in a total operating current requirement for the control and high-voltage device gate of less than 2.5 mA for a switching frequency of 100 kHz.

Concluding remarks

The new technology of power integrated circuits is making progress similar to that in VLSI circuitry in terms of size reduction and fabrication technology. The combination of control circuitry and power devices on the same silicon wafer provides significant opportunities for cost reduction and improved performance in power control applications. Applications

such as power supplies, electronic ballasts, domestic appliances, automotive switches, and others are expected to benefit as the new technology becomes assimilated into power systems design.

The valuable contributions to the development of PICs by the many members of the device, process, IC & systems design, CAD, and clean-room groups are gratefully acknowledged.

Summary. Power integrated circuits (PICs) are becoming an important branch of microelectronics. These circuits can be applied in motor control, power supplies, lighting and automotive areas. One specific category of PIC is known as the High-Voltage IC (HVIC). The key to the economically successful design of PICs is the ability to produce devices which can handle high voltages (several hundred volts) or high currents (up to tens of amperes) with a small silicon area. New device techniques which achieve this are discussed. Since many PICs have to be designed for specific applications (ASPICs), special design tools have been evolved which allow the rapid design ('silicon compilation') and simulation of the entire power system. Examples of the application of these tools to a switched-mode power supply are presented.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories, Sunnyvale, P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

A. G. Dirks & J. J. van den Broek	E	θ -Al ₂ Cu formation at room temperature in metastable Al-Cu alloy films	Acta Metall. 37	9-15	1989
G. N. A. van Veen, T. S. Baller & J. Dieleman	E	A time-of-flight study on the nanosecond laser induced etching of Cu with Cl ₂ at 308 nm	Appl. Phys. A 47	183-192	1988
P. Bellon*, J. P. Chevalier*, G. P. Martin* (*C.E.C.M.-C.N.R.S., Vitry), E. Dupont-Nivet, C. Thiebaut & J. P. André	L	Chemical ordering in Ga _x In _{1-x} P semiconductor alloy grown by metalorganic vapor phase epitaxy	Appl. Phys. Lett. 52	567-569	1988
P. J. A. Thijs, E. A. Montie, H. W. van Kesteren & G. W. 't Hooft	E	Atomic abruptness in InGaAsP/InP quantum well hetero-interfaces grown by low-pressure organometallic vapor phase epitaxy	Appl. Phys. Lett. 53	971-973	1988
E. A. Montie, P. J. A. Thijs & G. W. 't Hooft	E	Photoluminescence excitation spectroscopy of Ga _x In _{1-x} As _y P _{1-y} /InP	<i>ibid.</i>	1611-1613	1988
A. E. T. Kuiper, M. F. C. Willemsen, J. M. G. Bax & F. H. P. H. Habraken (Univ. Utrecht)	E	Oxidation behaviour of LPCVD silicon oxynitride films	Appl. Surf. Sci. 33-34	757-764	1988
A. Valster, J. P. André, E. Dupont-Nivet & G. M. Martin	E,L	High-power AlGaInP three-ridge type laser diode array	Electron. Lett. 24	326-327	1988
C. W. J. Beenakker, H. van Houten & B. J. van Wees (Univ. of Technol., Delft)	E	Mode interference effect in coherent electron focusing	Europhys. Lett. 7	359-364	1988
K. A. Schouhamer Immink	E	Coding techniques for partial-response channels	IEEE Trans. COM-36	1163-1165	1988
F. W. A. Dirne & M. Brouha	E	Soft-magnetic properties of microcrystalline Co-Fe-Si-B alloys prepared by sputtering	IEEE Trans. MAG-24	1862-1864	1988
P. Philippe, W. El-Kamali & V. Pauker	L	Physical equivalent circuit model for planar Schottky varactor diode	IEEE Trans. MTT-36	250-255	1988
F. E. J. Kruseman Aretz	E	On a recursive ascent parser	Inf. Process. Lett. 29	201-206	1988
N. Willems*, R. Collier* & J. 't Hart* (*Inst. for Perception Res., Eindhoven)		A synthesis scheme for British English intonation	J. Acoust. Soc. Am. 84	1250-1261	1988
E. W. Meijer, S. Nijhuis & F. C. B. M. van Vroonhoven	E	Poly-1,2-azepines by the photopolymerization of phenyl azides. Precursors for conducting polymer films	J. Am. Chem. Soc. 110	7209-7210	1988
G. N. A. van Veen, T. Baller* & A. E. de Vries* (*FOM, Amsterdam)	E	A time-of-flight study of the neutral species produced by nanosecond laser etching of CuCl at 308 nm	J. Appl. Phys. 60	3746-3749	1986

- P. Boher, M. Renaud, L. J. van IJzendoorn, J. Barrier & Y. Hily *E,L* Structural and electrical properties of silicon nitride films prepared by multipolar plasma-enhanced deposition *J. Appl. Phys.* **63** 1464-1472 1988
- G. J. van Gorp, D. L. A. Tjaden, G. M. Fontijn & P. R. Boudewijn *E* Zinc diffusion in InGaAsP *J. Appl. Phys.* **64** 3468-3471 1988
- A. H. van Ommen, A. H. Reader & J. W. C. de Vries *E* Influence of microstructure on the resistivity of MoSi₂ thin films *ibid.* 3574-3580 1988
- P. C. P. Bouten & G. de With *E* Crack nucleation at the surface of stressed fibers *ibid.* 3890-3900 1988
- A. J. M. Kaizer & A. Leeuwestein *E* Calculation of the sound radiation of a nonrigid loudspeaker diaphragm using the finite-element method *J. Audio Eng. Soc.* **36** 539-551 1988
- M. de Keijser, C. van Opdorp & C. Weber *E* Peculiar asymmetric flow pattern in a vertical axisymmetric VPE reactor *J. Cryst. Growth* **92** 33-36 1988
- H. F. J. van 't Blik & H. J. M. Boerrigter-Lammers *E* On the MOVPE growth of self-aligned laser structures *ibid.* 165-170 1988
- P. J. A. Thijs, T. van Dongen, P. I. Kuindersma, J. J. M. Binsma, L. F. Tiemeyer, J. M. Lagemaat, D. Moroni & W. Nijman *E,L* High quality InGaAsP-InP for multiple quantum well laser diodes grown by low-pressure OMVPE *J. Cryst. Growth* **93** 863-869 1988
- M. Erman, P. Jarry, R. Gamonal, J.-L. Gentner, P. Stephan & C. Guedon *L* Monolithic integration of a GaInAs p-i-n photodiode and an optical waveguide: modeling and realization using chloride vapor phase epitaxy *J. Lightwave Technol.* **LT-6** 399-412 1988
- U. Dressel*, U. Meister*, E. Dormann* (*Univ. Bayreuth) & K. H. J. Buschow *E* Nuclear quadrupole interaction of gadolinium and europium in cubic, ferromagnetically ordered intermetallic compounds *J. Magn. & Magn. Mater.* **74** 91-100 1988
- Y. L. Raikher (*Inst. of Continuous Media Mechanics, Perm*) & P. C. Scholten *E* Magnetic colloid in an ac magnetic field; constant birefringence effect *ibid.* 275-280 1988
- F. J. C. M. Toolenaar *E* Silica-induced exaggerated grain growth in MnZn ferrite *J. Mater. Sci.* **23** 3144-3150 1988
- S. M. Wolfrum *E* Dry pressing of surface-modified powders *J. Mater. Sci. Lett.* **7** 1130-1132 1988
- A. H. Boonstra & C. A. M. Mulder *E* Effect of hydrolytic polycondensation of tetraethoxysilane on specific surface area of SiO₂ gels *J. Non-Cryst. Solids* **105** 201-206 1988
- H. T. Hintzen & H. M. van Noort *E* Investigation of luminescent Eu-doped sesquioxides Ln₂O₃ (Ln = In, Sc, Y, La, Gd, Lu) and some mixed oxides by ¹⁵¹Eu Mössbauer spectroscopy *J. Phys. Chem. Solids* **49** 873-881 1988
- H. Verweij & W. H. M. Bruggink *E* Precision determination of the stoichiometry parameter *x* of YBa₂Cu₃O_{*x*} *ibid.* 1063-1069 1988
- R. Eppenga, M. F. H. Schuurmans & H. W. A. M. Rompa *E* *Ab-initio* determinations of semiconductor spin-orbit splittings from ASW *ibid.* 1119-1124 1988
- A. H. El-Sayed*, G. J. Nieuwenhuys*, J. A. Mydosh* (*Univ. Leiden) & K. H. J. Buschow *E* Electron spin resonance in ternary intermetallic compounds with MgAgAs structure *J. Phys. F* **18** 2265-2281 1988
- M. Erman, C. Alibert*, J. A. Cavaillès, P. Frijlink & C. Bouche* (*Univ. des Sciences et Techniques du Languedoc, Montpellier) *L* Assessment of a multiple quantum well and a superlattice structure by spectroscopic ellipsometry, electroreflectance and photorefectance modulation spectroscopy *J. Physique* **48** (Colloque C5) C5/139- C5/142 1987
- D. Moroni, J. N. Patillon, E. P. Menu, P. Gentric & J. P. André *L* Optical investigation of InGaAs-InP quantum wells *ibid.* C5/143- C5/146 1987
- J. Haisma *E* SOI technologies: their past, present and future *J. Physique* **49** (Colloque C4) C4/3- C4/12 1988
- A. J. Walker & P. H. Woerlee *E* A mobility model for MOSFET device simulation *ibid.* C4/265- C4/268 1988
- M. G. Pitt & P. A. van der Plas *E* Limitations on n⁺/p⁺ spacing due to shadowing effects in a 0.7 μm retrograde well CMOS process *ibid.* C4/553- C4/556 1988
- P. T. J. Biermans, T. Poorter & H. J. H. Merks-Eppingbroek *E* The impact of different hot-carrier-degradation components on the optimization of submicron n-channel LDD transistors *ibid.* C4/787- C4/790 1988

- | | | | | |
|---|--|------------------------------------|-------------|------|
| M. L. Verheijke & R. M. W. Jansen
E | The single-comparator method in thermal neutron activation analysis extended for some (n,p) reactions | J. Radional. & Nucl. Chem. 125 | 103-111 | 1988 |
| M. L. Verheijke & R. M. W. Jansen
E | Determination of characteristic neutron flux parameters of nuclear reactors for thermal neutron activation analysis | <i>ibid.</i> | 113-125 | 1988 |
| T. S. Baller, G. N. A. van Veen & J. Dieleman
E | The role of chlorinated surface films in excimer laser etching of Cu at low Cl ₂ pressures | J. Vac. Sci. & Technol. A 6 | 1409-1413 | 1988 |
| A. Heerschap, J. J. van Vaals, A. H. Bergman, J. H. den Boef, P. H. J. van Gerwen & E. J. 's-Gravenmade (Univ. Groningen)
E | <i>In vivo</i> ³¹ P NMR studies of rat salivary glands at 6.3 Tesla | Magn. Resonance in Med. 8 | 129-141 | 1988 |
| A. H. van Ommen
E | Low dislocation SOI by oxygen implantation | Mater. Res. Soc. Symp. Proc. 107 | 43-52 | 1988 |
| J. Haisma, C. Langereis, J. M. M. Pasmans & J. E. J. Schmitz
E | Laser recrystallization of polysilicon on mono-crystalline insulating substrates | Microelectron. Eng. 8 | 105-120 | 1988 |
| A. Djemel*, J. Castaing* (* Lab. de Physique des Matériaux, CNRS, Meudon) & M. Duseaux
L | Interaction between dislocations and In in In-doped GaAs single crystals under high-temperature plastic deformation | Philos. Mag. A 57 | 671-676 | 1988 |
| R. Piotrkowski*, J. L. Robert* (*Univ. des Sciences et Techniques du Languedoc, Montpellier), E. Litwin-Staszewska (Polish Academy of Sciences, Warsaw) & J. P. André
L | Pressure study of metastability in Ga _{1-x} Al _x As/GaAs:Si heterojunctions | Phys. Rev. B 37 | 1031-1034 | 1988 |
| G. L. J. A. Rikken*, J. A. M. M. van Haaren*, W. van der Wel (Univ. of Technol., Delft), A. P. van Gelder+, H. van Kempen+ (* Univ. Nijmegen), P. Wyder* (* Max-Planck-Inst. für Festkörperforschung, Grenoble & Univ. Nijmegen), J. P. André, K. Ploog (Max-Planck-Inst. für Festkörperforschung, Stuttgart) & G. Weimann (Forschungsinst. der Deutschen Bundespost, Darmstadt)
L | Two-terminal resistance of quantum Hall devices | <i>ibid.</i> | 6181-6186 | 1988 |
| R. Eppenga & M. F. H. Schuurmans
E | Effect of bulk inversion asymmetry on [001], [110], and [111] GaAs/AlAs quantum wells | <i>ibid.</i> | 10923-10926 | 1988 |
| C. W. J. Beenakker & H. van Houten
E | Boundary scattering and weak localization of electrons in a magnetic field | Phys. Rev. B 38 | 3232-3240 | 1988 |
| F. P. Widdershoven
E | Ionized-impurity scattering due to clusters of correlated impurities | <i>ibid.</i> | 3391-3394 | 1988 |
| J. Aarts, A.-J. Hoeven & P. K. Larsen
E | Core-level study of the phase transition on the Ge(111)-c (2x8) surface | <i>ibid.</i> | 3925-3930 | 1988 |
| G. E. W. Bauer & T. Ando (Univ. Tokyo)
E | Exciton mixing in quantum wells | <i>ibid.</i> | 6015-6030 | 1988 |
| L. W. Molenkamp, G. E. W. Bauer, R. Eppenga & C. T. Foxon
E,R | Exciton binding energy in (Al,Ga)As quantum wells: effects of crystal orientation and envelope-function symmetry | <i>ibid.</i> | 6147-6150 | 1988 |
| P. Kuiper*, G. Kruizinga*, J. Ghijsen*, M. Grioni*, P. J. W. Weijs*, F. M. F. de Groot*, G. A. Sawatzky* (*Univ. Groningen), H. Verweij, L. F. Feiner & H. Petersen (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung, Berlin)
E | X-ray absorption study of the O 2p hole concentration dependence on O stoichiometry in YBa ₂ Cu ₃ O _x | <i>ibid.</i> | 6483-6489 | 1988 |
| P. Astié*, J. J. Couderc*, P. Chomel*, D. Quelard* (* Lab. de Physique des Solides, Toulouse) & M. Duseaux
L | Thermal activation of plastic deformation of undoped GaAs between 528 and 813 K | Phys. Stat. Sol. a 96 | 225-242 | 1986 |
| M. G. J. Heijman
E | Reactive sputter etching of magnetic materials in an HCl plasma | Plasma Chem. & Plasma Processing 8 | 383-397 | 1988 |
| H. M. J. Boots & N. A. Dotson (Univ. of Minnesota, Minneapolis, MN)
E | The simulation of free-radical cross-linking polymerization: the effect of diffusion | Polym. Commun. 29 | 346-349 | 1988 |

L. Minnema & J. M. van der Zande	E	Pattern generation of polyimide coatings and its application in an electrophoretic image display	Polym. Eng. & Sci. 28	815-822	1988
L. M. G. Feijs	E	A formalisation of design structures	Proc. COMP EURO 88 — System Design: Concepts, Methods and Tools, Brussels 1988	214-229	1988
E. Delhay, C. Rocher, M. Fichelson & I. Lecuru	L	A 3.0 ns, 330 mW 8x8 GaAs Booth's multiplier	Proc. IEEE Gallium Arsenide Integrated Circuit Symp., Portland, OR, 1987	249-252	1987
L. J. M. Esser, H. C. G. van Kuijk, J. G. C. Bakker, C. H. L. Weijtens & A. J. P. Theuwissen	E	A smearfree accordion CCD imager	Proc. ICCE Int. Conf. on Consumer electronics, Rosemont, IL, 1988	10-11	1988
P. J. Severin & W. Bardeel	E	Evanescence-mode coupling in a bicore fibre: experiments and applications	Proc. SPIE 949	220-228	1988
P. Pesqué & C. Blanc	L	Increasing of the grating lobe effect in multiscatters medium	Proc. Ultrasonics Symp., Vol. 2, Denver, CO, 1987	849-852	1987
L. J. M. Kuijpers	E	Impact of the decrease in CFC emissions on refrigeration: target of the IIR initiative	Rev. Int. Froid 11	371-384	1988
M. Jimenez-Melendo*, A. Djemel*, J. P. Rivière*, J. Castaing* (*Lab. de Physique de Matériaux, CNRS, Meudon), C. Thomas & M. Duseaux	L	Influence of indium on the dissociation of dislocations in GaAs of high temperature	Rev. Phys. Appl. 23	251-255	1988
S. Makram-Ebeid & P. Boher	L	Defect pairs and clusters related to the EL2 centre in GaAs	<i>ibid.</i>	847-862	1988
P. F. Fontein*, J. M. Lagemaat, J. Wolter* (*Univ. of Technol., Eindhoven) & J. P. André	E,L	Magnetic field modulation — a method for measuring the Hall conductance with a Corbino disc	Semicond. Sci. Technol. 3	915-918	1988
R. H. Coursant, J. M. Tellier, L. Eyraud*, P. Eyraud* (*INSA, Lab. de Génie Electrique et Ferroélectricité, Villeurbanne) & M. Fink (Univ. de Paris VII, Paris)	L	Characterization of modified lead titanate piezoceramics. Application to the design of array transducers	Sensors & Actuators 13	351-363	1988
C. Mequio, R. H. Coursant & J. M. Tellier	L	Characterization of piezoplastics	Sensors & Actuators 14	1-8	1988
M. Klee, G. M. Stollman, S. Stotz & J. W. C. de Vries	E,A	Preparation and transport properties of superconducting layers in the Ca-Sr-Bi-Cu-O system	Solid State Commun. 67	613-617	1988
A. H. van Ommen	E	Diffusion of ion-implanted group III and V impurities in SiO ₂	Solid State Phenomena 1&2	133-152	1988
P. W. J. M. Boumans, J. J. A. M. Vrakking & A. H. M. Heijms	E	Mutual spectral interferences of rare earth elements in inductively coupled plasma atomic emission spectrometry — II. Approach to the compilation and use of pseudo physically resolved spectral data	Spectrochim. Acta 43B	1365-1404	1988
P. America & J. de Bakker (Centre for Mathematics and Computer Science, Amsterdam)	E	Designing equivalent semantic models for process creation	Theor. Comput. Sci. 60	109-176	1988
J. J. P. Bruines	L	Time-resolved study of solidification phenomena on pulsed-laser annealing of amorphous silicon	Thesis, Eindhoven	1-175	1988
J. J. M. Ruigrok	E	Theoretical and experimental investigations on short-wavelength recording	Thesis, Enschede	1-554	1988
P. Boher, J. Schneider, M. Renaud & J.-P. Landesman	L	Passivation des semiconducteurs III-V: utilisation de plasmas multipolaires en ambiance ultraviolette	Vide/Couches Minces 43	207-213	1988
R. Mabon, J. P. Landesman, B. J. Schäfer & J. Bonnet (Univ. de Paris Sud, Orsay)	L	Spectrométrie de photoelectrons X et UV pour le suivi <i>in-situ</i> de l'exposition de GaAs (100) et (110) a des plasmas multipolaires	<i>ibid.</i>	251-252	1988

Contents

A farewell message K. Bulthuis	325
The Compact Disc Interactive system B. A. G. van Luyt and L. E. Zegers <i>Information combined as pictures, sound and text can be located and retrieved in a user/system dialogue</i>	326
Then and Now (1939-1989)	334
Phosphor screens in cathode-ray tubes for projection television R. Raue, A. T. Vink and T. Welker <i>Phosphor screens for projection television are very different from screens for direct-view television</i>	335
Electron guns for projection television T. G. Spanjer, A. A. van Gorkum and W. M. van Alphen <i>A new electron-optical design gives the bright and sharp pictures required for high-definition projection television</i>	348
A diagnostic X-ray tube with spiral-groove bearings E. A. Muijderman, C. D. Roelandse, A. Vetter and P. Schreiber <i>Spiral-groove bearings with liquid metal as a lubricant conduct heat away and pass electric current</i>	357
Scientific publications	364
Subject index, Volumes 36-44	365
Author index, Volumes 36-44	374

P. Geittner and H. Lydtin, Manufacturing optical fibres by the PCVD process,

Philips Tech. Rev. 44, No. 8/9/10, 241-249, May 1989.

Optical fibres for telecommunications are either multimode fibres (step- or graded-index) or single-mode fibres. Single-mode fibres have either a stepped profile with a small core diameter or a complex layered refractive-index profile. The last type of fibres are called DFMS (Dispersion-Flattened Single-Mode) fibres. In fibres of this type pulse broadening due to mode dispersion does not occur and material dispersion and waveguide dispersion almost completely compensate one another. The Philips PCVD (Plasma-activated Chemical Vapour Deposition) process can be used for the manufacture of fibres with complicated index profiles. Closely controlled amounts of dopants such as GeO_2 can be introduced into the basic SiO_2 material at very low concentrations. The concentration of OH^- ions, which can introduce an intolerably high attenuation, has been greatly reduced by adding C_2F_6 to the reactive gas mixture. This reduced the attenuation due to OH^- ions to a minimum value of 0.1 dB/km.

D. J. Gravesteijn, C. J. van der Poel, P. M. L. O. Scholte and C. M. J. van Uijen, Phase-change optical recording,

Philips Tech. Rev. 44, No. 8/9/10, 250-258, May 1989.

In phase-change optical recording the recording layer of a disc is heated locally by a focused laser beam. Heating crystalline material to just above the melting point and then cooling it rapidly causes it to change from the crystalline phase to the amorphous phase, while heating amorphous material below the melting point has just the opposite effect. The read-out of written effects depends on the difference in optical properties between the two phases. The choice of materials for the recording layer is determined not only by the optical properties, but also by the crystallization behaviour. The stability of amorphous areas and the crystallization rate at higher temperatures are particularly important here. A material with a very high crystallization rate, GaSb, is found to be suitable for non-erasable recording in an amorphous recording layer. For erasable recording in a crystalline recording layer all the requirements can be met with an InSbTe alloy. The first results of recording experiments with these materials look very promising.

O. Dössel, M. H. Kuhn and H. Weiss, Magnetic fields in medical diagnostics: MR and SQUID,

Philips Tech. Rev. 44, No. 8/9/10, 259-267, May 1989.

A magnetic-resonance (MR) system with a two-tesla superconducting magnet was installed at the Philips laboratories in Hamburg in 1983. Good images of cross-sections of the human body were produced at this high field, but there were difficulties due to the chemical-shift effect. This effect gave a blurred image because the images due to water and fat did not coincide. The difficulty was resolved by suppressing one of these images. This was a first step in the development of MR spectrometry, an analytical method that gives information about the chemical processes in a small region of the human body. As the main field increases the resolution of MR spectrometry improves. An MR system with the high flux density of four teslas was therefore developed and brought into operation. Another development from the Hamburg laboratories is a SQUID magnetometer (SQUID stands for Superconducting QUantum Interference Device), which will be used for detecting the extremely weak magnetic field generated by the currents in the neurons of the brain. A SQUID consists of a superconducting ring with two Josephson junctions. The SQUID will be used in reconstructing an image of neuronal activity from measurements of the field around the brain.

P. Blood, C. T. Foxon and E. D. Fletcher, The application of semiconductor superlattices to short-wavelength lasers,
Philips Tech. Rev. 44, No. 8/9/10, 268-273, May 1989.

This research shows that short-period superlattice structures can be used to advantage in a variety of short-wavelength semiconductor lasers. Molecular beam epitaxy has been used to grow laser structures in which the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy regions have been replaced by all-binary short-period $(\text{AlAs})_n(\text{GaAs})_m$ superlattices embodying binary layers as thin as three monolayers. Quantum-well lasers with superlattice barriers and GaAs wells have a lower threshold current than equivalent structures using alloys. An analysis of the performance of devices with different cavity lengths shows that the optical-scattering losses in these structures are indeed very low. There is evidence for a leakage current associated with recombination in the barrier regions though it should be possible to reduce this by modifying the design of the superlattice barrier. Double-heterostructure lasers have been made with an all-binary superlattice active region which operates at 786 nm and with encouraging values of threshold current.

M. F. H. Schuurmans, R. Coehoorn, R. Eppenga and P. J. Kelly, Predicting the properties of materials: dream or reality?,
Philips Tech. Rev. 44, No. 8/9/10, 276-286, May 1989.

Before new materials are used in commercial products their properties must be known. This information can be acquired from experiments, of course. Properties can also be calculated even before the materials exist. One of the theories that can be used in these calculations is the density functional theory. The present developments in this field are discussed with the aid of results obtained from the theory. These results relate to materials in the ground state: the different crystal structures of silicon at increasing pressure, the rearrangement of atoms at the surface of silicon, the consequences for the crystal lattice when oxygen is incorporated and the magnetic behaviour of iron-yttrium compounds. The theory can also be used, with some modification, for predicting the emission of light from GaAs/AlAs superlattices.

A. Bruffaerts, E. Henin and A. Pirotte, A sound basis for the generation of explanations in Expert Systems,
Philips Tech. Rev. 44, No. 8/9/10, 287-295, May 1989.

In this article we have argued that a wide spectrum of explanations can be useful in an expert system. We have looked briefly at attempts to generate some of these explanations. A general architecture has been proposed for expert systems, clearly distinguishing between the functions of two cooperating agents, a dialogue manager and a problem-solver. We have argued that a logical framework is most convenient for a precise declarative version of the knowledge base and of its relationships to query answers from the problem-solver. Our own work is aimed at the development of a powerful problem-solver component along those lines. Formalisms and tools have been developed for generating a form of proof tree that offers a sound basis for the conventional trace-based how-, why-, why-not-explanations. A logic-based object-oriented knowledge-representation formalism is being designed to support a uniform encoding of knowledge and meta-knowledge for the generation of more conceptual explanations.

G. Conner and R. H. Lane, HS3: an advanced bipolar-IC technology,

Philips Tech. Rev. 44, No. 8/9/10, 296-301, May 1989.

HS3 has been shown to be a versatile, high-performance IC-manufacturing process. The process consists of a single core process and a number of modular additions. Each addition requires an extra masking step and corresponds to an optional product family, such as analog circuits, ECL gate arrays, memory and logic circuits, and operational amplifiers for automotive applications. High production numbers in the core process facilitate statistical process control. The excellent properties of npn transistors with 1 by 3 μm emitter are demonstrated by a Gummel plot and a plot of cut-off frequency f_T against collector current.

M. Rocchi, Research on monolithic GaAs MESFET circuits at LEP,

Philips Tech. Rev. 44, No. 8/9/10, 302-309, May 1989.

In GaAs MESFETs (MEtal-Semiconductor Field-Effect Transistors), the gate metal is directly applied to the GaAs surface to form a Schottky-barrier diode. The transit time of electrons beneath the gate electrode can be very short (down to a few ps) so that these transistors have very good high-frequency (> 1 GHz) characteristics. During the last fifteen years monolithic circuits based on GaAs MESFETs have been extensively investigated at LEP. Significant progress in device technology has greatly widened the field for both digital and analog applications. High-speed low-power digital GaAs ICs can be used in instrumentation and telecommunication systems, in supercomputers and in systems for fast signal processing. Analog GaAs ICs can be used in various r.f. and microwave devices, not only for professional applications at frequencies up to 30 GHz, but also for consumer applications such as the 12-GHz front-ends for direct satellite television reception. The future of GaAs MESFET ICs will depend closely on the evolution of special semiconductor heterostructures for high-frequency devices and also on the economic effects of the submicron miniaturization of GaAs devices and their silicon counterparts.

M. Amato, G. Bruning, S. Mukherjee and I. T. Wacyk, Power integrated circuits,

Philips Tech. Rev. 44, No. 8/9/10, 310-320, May 1989.

Power integrated circuits (PICs) are becoming an important branch of microelectronics. These circuits can be applied in motor control, power supplies, lighting and automotive areas. One specific category of PIC is known as the High-Voltage IC (HVIC). The key to the economically successful design of PICs is the ability to produce devices which can handle high voltages (several hundred volts) or high currents (up to tens of amperes) with a small silicon area. New device techniques which achieve this are discussed. Since many PICs have to be designed for specific applications (ASPICs), special design tools have been evolved which allow the rapid design ('silicon compilation') and simulation of the entire power system. Examples of the application of these tools to a switched-mode power supply are presented.

PHILIPS TECHNICAL REVIEW

Volume 44, No. 11/12, November 1989

A FAREWELL MESSAGE

After very careful consideration we have decided to discontinue the publication of Philips Technical Review. We should explain why we are taking this step.

From its earliest days Philips Technical Review set out to be something special, not just a popular descriptive magazine or another professional scientific journal. The editors' aim has always been to present the material in the clearest possible way, with well-written text and a generous provision of illustrations and graphics. The style is not too academic, and mathematical treatment is kept to a minimum without oversimplifying so far that the treatment becomes vague and superficial.

The great feature of the articles in Philips Technical Review, as compared with those in the professional scientific literature, is that each article in the Review is complete in itself. This means that the subject has to be treated in a wider context, and that the problems arising also have to be described before revealing the solution.

During its fifty years and more the objectives of Philips Technical Review have not changed. But the world has changed, and so too has research. We live in days of increasing specialization. More and more research is done in project groups, often working in cooperation with other companies or in a European context, with team members from a range of disciplines. Describing a project properly — to the high standards of this journal — becomes a more complicated activity, and takes longer, so that the published version may no longer be current. Readers also have to work harder to follow the details of

the new developments in a rapidly expanding range of fields.

The changes in research have been matched by changes in the methods of disseminating information. There are more publications with a popular scientific content, and radio and television now have much to offer. We began to wonder whether such a journal was really the best way of presenting news about Philips research.

Considerations such as these led us to the decision to discontinue Philips Technical Review. This was no easy matter, since we have always been rather proud of the Review. And we do realize that we shall disappoint a large number of faithful readers, both inside Philips and outside.

So this is the final issue of Philips Technical Review, an issue strongly oriented toward the future. It discusses the Compact Disc Interactive system, research on high-definition projection television, and a new application of spiral-groove bearings. These subjects show that Philips research and development occupy a leading position in the world, and we firmly believe that the media will continue to keep you informed of our progress.

Thank you for your support and for the interest you have shown in Philips Technical Review.

K. BULTHUIS
Senior Managing Director of Philips Research

The Compact Disc Interactive system

B. A. G. van Luyt and L. E. Zegers

Two of the terms always associated with the Compact Disc Digital Audio system (CD-DA) are 'digital' and 'laser'. When the system was first introduced in the early eighties it started a real revolution in sound reproduction. In 1987 more than 30 million CD players and 450 million discs were sold. One of the systems derived from CD-DA is CD-I, which has two more special terms associated with it, 'interactive' and 'multi-media', since this system combines images, sound, text and software in an active dialogue with the user.

Introduction

The beam of light emitted by a laser can be focused to an extremely small spot, with a diameter of about a micron. This led to the idea, in the early seventies at Philips Research Laboratories, of using a laser for the recording and playback of information. The extensive research that followed eventually resulted in two new systems:

- LaserVision, for the recording and playback of video information in *analog* form^[1].
- Compact Disc Digital Audio (CD-DA), for the recording and playback of audio information in *digital* form^[2].

It may help if we briefly consider the similarities and differences between the LaserVision and Compact Disc systems. The most important similarity is that in both systems the signal is recorded on the disc in a long spiral track consisting of a succession of pits about 0.5 μm wide; see *fig. 1a*. The regions between the pits are called 'lands'. The pits are impressed into a plastic substrate by a mould or 'stamper', which is a pressing from a 'mother disc', which in turn is a copy of a master disc. The master disc is the result of 'burning' the information into a 'virgin' disc by a laser. In both systems the pits in the substrate are

protected by a transparent layer. Dust and surface damage cannot appear in the focal plane of the 'optical pick-up', which reads the information on the disc in the player.

An essential difference between the two systems is that analog recording is used for LaserVision and digital recording for Compact Disc. *Fig. 1b* shows how the analog signal is recorded on a LaserVision disc. An analog sound signal is superimposed on the frequency-modulated video signal. The resulting signal is limited in both the positive and negative directions. The leading and trailing edges of the blocks produced in this way form the pit/land transitions on the disc. In the CD-Video system derived from LaserVision the sound signal superimposed on the video signal is not an analog signal, but a digital signal.

Fig. 1c shows how the digital signal, a sequence of the values '0' and '1', is recorded on a Compact Disc. The length of each pit or land is always a multiple of 0.3 μm . This is different from a LaserVision disc, in which the pit lengths can have an infinite number of values. Every transition from pit to land or vice versa on a Compact Disc forms a 'bit' of value 1. The intermediate bits, which correspond to distances of 0.3 μm on the disc, have the value 0. These 'channel bits' have been produced by coding the original signal bits. These in turn have been produced from a succession

Ir B. A. G. van Luyt is with American Interactive Media Inc., Los Angeles, California, U.S.A., and was formerly with the Philips Consumer Electronics Division in Eindhoven. Dr Ir L. E. Zegers (Deputy Director) is with the Philips Consumer Electronics Division in Eindhoven.

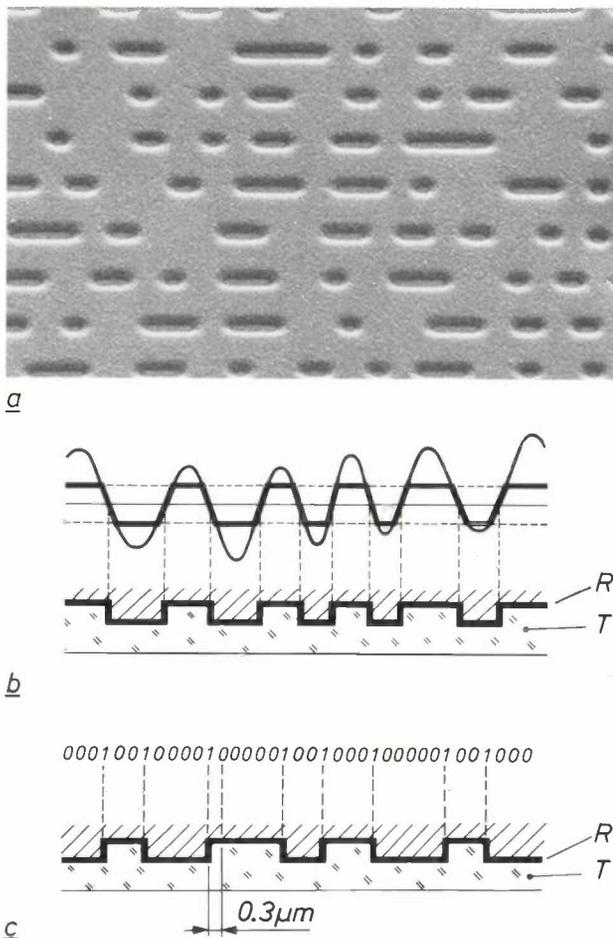


Fig. 1. *a)* The pits and 'lands' (the regions between the pits) in a disc for LaserVision or Compact Disc Digital Audio. The width of the pits is about $0.5\ \mu\text{m}$. The spiral of pits forms the 'track'. *b)* The conversion of an analog video signal into a sequence of pits and lands in LaserVision. Above: signal; below: cross-section of the disc. The frequency-modulated signal is limited in both positive and negative directions. The leading and trailing edges determine the location of the land/pit and pit/land transitions. *c)* The conversion of a digital audio signal into a sequence of pits and lands in CD-DA. The length of a pit or land is always a multiple of $0.3\ \mu\text{m}$. Above: bit sequence; below: cross-section of the disc. Every pit/land or land/pit transition corresponds to a '1' in the digital signal. In between the signal always has the value '0' for every $0.3\ \mu\text{m}$ of distance along the track. *R* reflecting layer. *T* transparent material.

of binary numbers that are sampled values of the original analog sound signal: pulse-code modulation or PCM. In the player a decoder circuit converts the channel bits into signal bits, and a digital-to-analog converter converts the signal bits into the analog sound signal.

The great advantage of the digital recording and reproduction of analog information is that it is insensitive to interference and noise. Also, reading errors due to damage or dirt at the surface of the disc can be corrected. The methods for signal processing, and many important quantities such as the dimensions of

the disc, have been specified in a CD-DA system standard drawn up by Philips and Sony. Licensing agreements have been concluded with many other companies.

After the successful introduction of the CD-DA system it soon became clear that a Compact Disc was not only exceptionally suitable for recording sound, but could be just as useful for storing digital information for computers. Talks with Sony resulted in a standard for a new kind of memory: CD-ROM (Compact Disc Read-Only Memory). This standard specifies that the digital information shall be organized in blocks, each with an address. Information in one or more blocks can then be traced rapidly and read out from the disc.

The CD-ROM disc is mainly intended for professional and business applications with personal computers. A CD-ROM disc with a diameter of 12 cm can contain about 650 megabytes of digital information. (1 megabyte is $2^{20} \times 8 = 1024 \times 1024 \times 8$ bits.) A disc can contain large numbers of names and addresses or other kinds of text, with a maximum storage capacity equivalent to 150 000 typed A4 pages.

The agreements for the CD-ROM disc have not reached the stage at which the interchangeability of disc and player is guaranteed at all times, as it is for CD-DA. A CD-ROM player is therefore usually a peripheral used with a particular make of computer. A group of companies known as the High Sierra Group have made a number of supplementary agreements relating to the organization of the information on a CD-ROM. These agreements have resulted in uniformity for the tables of contents linking data and addresses. The agreements are set out in ISO standard ISO 9660.

Philips and Sony have gone a step further by drawing up a common standard for a new data-storage system for *consumer application*: the Compact Disc Interactive system, or CD-I. This standard specifies a system and disc that will provide text, images and sound in a real-time dialogue with the user. A system that can offer such an extended range of information is called a 'multi-media environment'. The standard is sufficiently comprehensive to ensure that the disc can be played anywhere at any time.

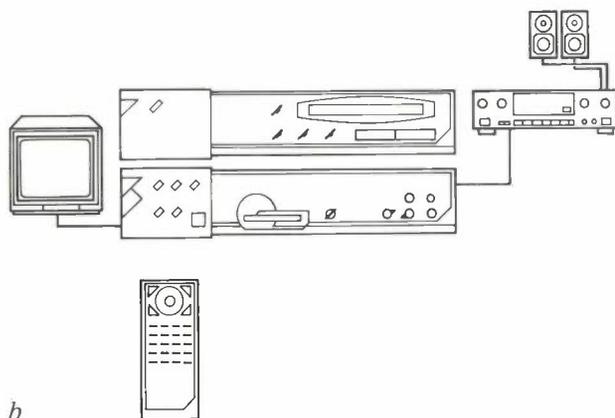
The agreements embodied in the CD-I standard do not only specify the organization, coding and processing of the data; they also specify the hardware. The CD-I player, which is virtually identical to a CD-DA player, must be connected to an MMC module (MMC

[1] The LaserVision system was originally called 'VLP' (for Video Long-Play); see Philips Tech. Rev. 33, 177-193, 1973.

[2] 'Compact Disc Digital Audio', Philips Tech. Rev. 40, 149-180, 1982.



a



b

Fig. 2. *a*) The CD-I player and the MMC module (Multi-Media Controller). *b*) Block diagram of the hardware. The MMC module decodes the signals for the user's television receiver and audio equipment. The user can communicate interactively with the system, e.g. with a remote control.

stands for Multi-Media Controller), see *fig. 2*. The MMC module contains processors for the video and audio signals and a microprocessor for data management. The module decodes the sound and image information and passes it to outputs connected to the user's television receiver and audio equipment. The data management comes under CD-RTOS (Compact Disc Real-Time Operating System), a special system derived from the OS-9 operating system.

The user gives instructions to his CD-I system by moving a cursor on the screen. He controls the cursor with a controller such as a 'mouse' or a 'joystick'. *Fig. 3* gives an example of the interactive use of an experimental CD-I with the menus shown on the screen. This CD-I was specially made for demonstration purposes.

The CD-I system is the result of extensive experience in optical recording and the interactive use of information systems, gathered from various parts of our company. Here we should mention the experience

obtained in compiling an electronic dictionary. This was a joint project shared between our colleagues at CTI (Centre de Technologie Informatique, a Philips company) in Paris and Philips Research Laboratories at Redhill, England^[3]. Knowledge already existing within the company about computer operating systems such as OS-9 has been very useful here.

The first working models of CD-I hardware were constructed by the Predevelopment department of the former Home Interactive Systems group (now Interactive Media Systems), in 1985. These models were used in the joint efforts with Sony to establish a system standard. They were also used to specify the requirements for the integrated circuits in VLSI technology (Very-Large-Scale Integration).

Derived versions of these models have also been used as 'authoring systems': 'tools' for suppliers of software for interactive programmes. Authoring systems are necessary for classifying and coding the image, sound and text information. The processed information is then permanently recorded in the stamper for the discs. In the meantime the first CD-Is for demonstration purposes had become available, see for example *fig. 3*. A first test batch of CD-I hardware was also ready in late 1988.

In the rest of the article we shall first look more closely at the standards for CD-DA, CD-ROM and CD-I. Then we shall discuss CD-I in rather more detail, with a look at the audio and video units and the CD-RTOS control system. Finally, we shall consider future developments.

The standards

Several standards have now been produced:

- The 'Red Book', for CD-DA (1982),
- The 'Yellow Book', for CD-ROM (1985), and
- The 'Green Book', for CD-I (1988).

The Yellow Book and the Green Book are augmented versions of the earlier standards.

In the *Red Book*, blocks of bits resulting from the sampling of an audio signal have blocks of 'parity bits' added to them, according to the rules for the Cross-Interleaved Reed-Solomon Code (CIRC). The blocks of parity bits allow a wide range of errors to be detected and corrected. The data stream is then modulated in Eight-to-Fourteen Modulation (EFM): blocks of eight bits are translated into blocks of fourteen channel bits. The requirement that must be satis-

[3] Valuable contributions to the architecture of CD-I were made by R. Bruno and E. Schylander (Interactive Media Systems Group, formerly known as Home Interactive Systems), J. Tailade (CTI) and S. R. Turner (PRL). Many others contributed to the design of CD-I, including J. Veldhuis (Interactive Media Systems), who created the first CD-I programs.

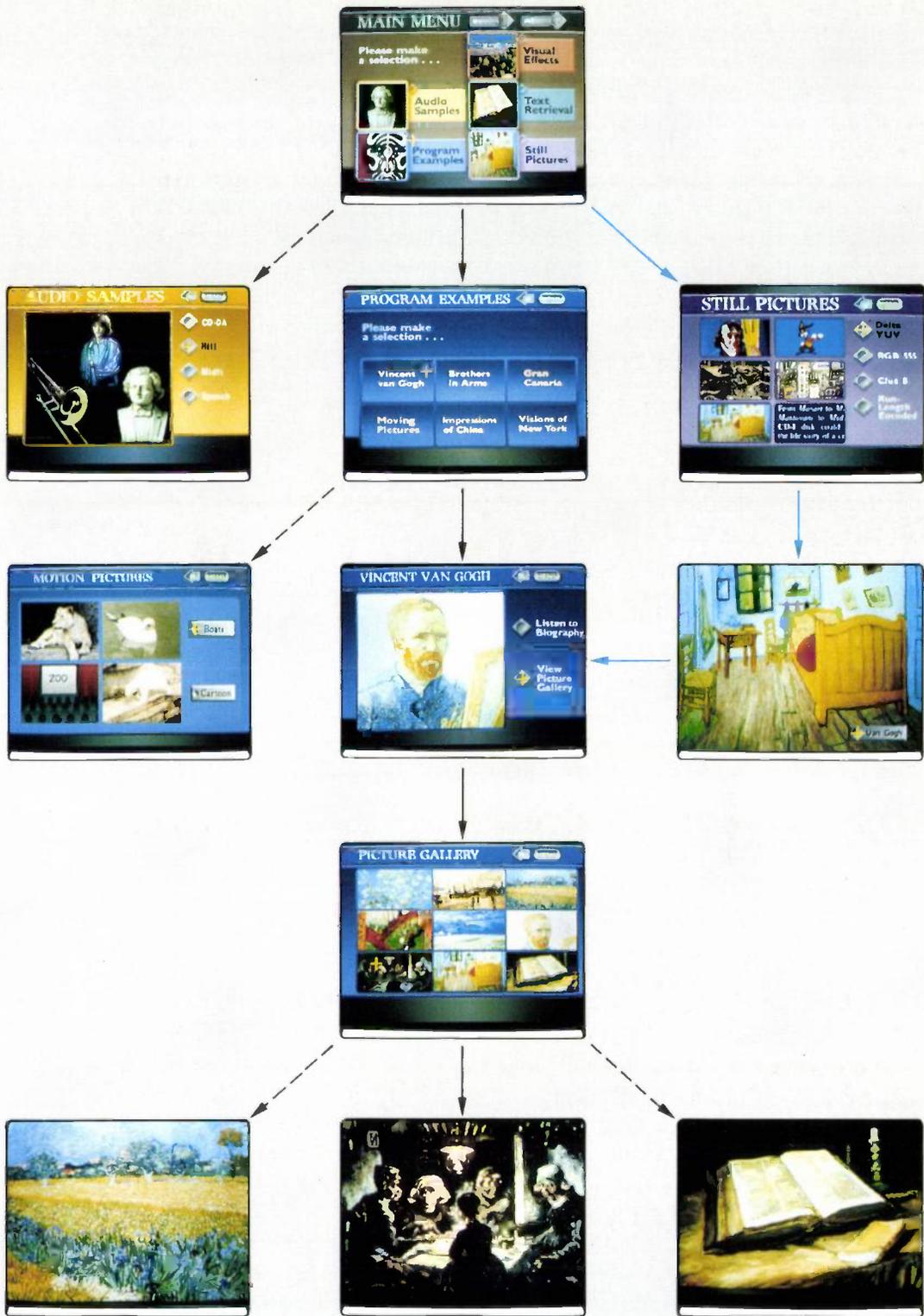


Fig. 3. Example of a CD-I dialogue. The user can find his way about the information on the disc by selecting from menus. The experimental CD-I here contains various examples of CD-I programs and was used for a first demonstration of the system. (The pictures shown here are not representative of the pictures from production CD-Is.) The black arrows correspond to the selections made by the user, as can be seen from the cursor positions. The blue arrows give an example of interactive use: after choosing 'STILL PICTURES' the user has selected a picture by Van Gogh, and has then gone on to obtain information about the painter by selecting 'VAN GOGH'. After the user has selected 'PICTURE GALLERY' and then one of the pictures, it appears on the screen and a reading of a translation of the corresponding letter from the painter's brother Theo is heard from the loudspeaker.

fied here is that sequences of zeros in the resulting stream of channel bits, see fig. 1c, must contain a minimum of two zeros and a maximum of ten, always separated by a '1'. Every sequence of zeros plus a '1' corresponds to a pit or land on the disc.

In the *Yellow Book*, the stream of channel bits on the disc is distributed among 'sectors', each containing 2352 bytes of the original information. Each sector starts with the same pattern of synchronization bits to identify the start of a sector. This is followed by a bit pattern representing the sector address and a bit pattern indicating the 'mode'. The sector addresses can be used in a contents list, which can be included at the start of every disc.

The *Yellow Book* also gives rules for mode designation. In Mode 1 more errors can be corrected than in CD-DA, since each sector contains 288 extra parity bits. This additional provision for error correction is necessary with computer data, and it ensures that no more than one error in a hundred million discs remains uncorrected. Mode 2 does not necessarily have this extra error correction, and is used for storing information in which the consequences of a very occasional error are less serious, such as sound and image information.

The *Green Book* defines the rules for the hardware, system software and the audio and video information in CD-I. These rules ensure that a disc can be used in any CD-I player. The rules for the organization of the information on a disc are mainly based on the *Yellow Book*. The function of the software is to present interleaved audio, video and text information in real-time dialogue with the user.

The *Green Book* specifies that in addition to the actual 'header' with the sector address each sector shall contain a 'subheader'. This subheader consists of four bytes, duplicated for extra reliability, and contains information about the type, format and quality level of the data in the sector. One of the functions of the subheader is to permit the real-time presentation of the information on the disc.

A distinction is made in the *Green Book* between the formats 'Form 1' and 'Form 2'. These offer much the same possibilities for error correction as Modes 1 and 2 in CD-ROM. Since the format designation is included in the subheaders in CD-I, sectors of different format can be interleaved on the disc. Form 1 is used for video information and computer data, Form 2 for

audio information and also for video information. A Form-2 sector can contain more information because it does not have the extra parity bits.

Characteristics and applications of CD-I

CD-I has been designed for a multiplicity of applications. These can be subdivided into the following main groups:

- education and training, e.g. language courses, encyclopaedias and 'talking books';
- entertainment, e.g. adventure games and other kinds of interactive games;
- creative leisure, e.g. drawing, painting and composing;
- touring and traffic. This includes consulting maps and tracing out routes. The CARIN vehicle navigation system (CARIN stands for Car Information and Navigation system) makes use of CD-I^[4].

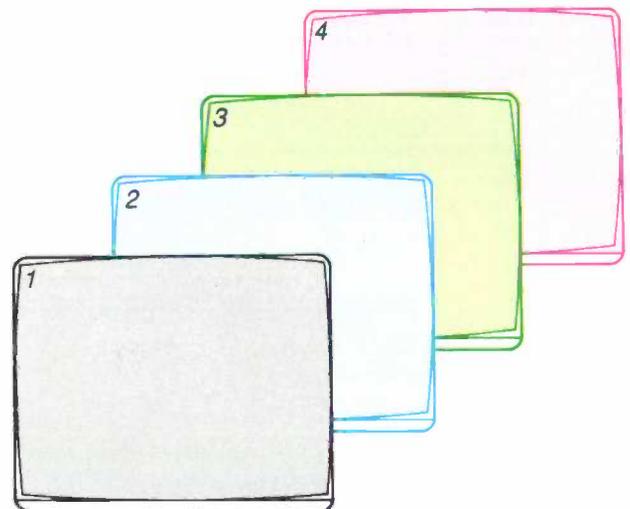


Fig. 4. The four image planes in CD-I corresponding to the four image memories. Image plane 4 can be the background. The image planes can be combined on the screen; when this is done a higher-level image is suppressed, so that it becomes 'transparent'. A lower-level image can then be seen; see also figs 5 and 6.

The organization, digitization and coding of images, sound and text and the provision of paths for interactive use is a time-consuming creative process that requires the use of an 'authoring system'. The result of such a process is a large quantity of digital information on a conventional magnetic recording medium. This is used in making the mother disc in a CD factory.

Companies that are going to supply CD-I programs already have authoring systems. These consist of a CD-I player and MMC module, with extra software

^[4] M. L. G. Thoone, CARIN, a car information and navigation system, Philips Tech. Rev. 43, 317-329, 1987.

^[5] R. J. Sluyter, Digitization of speech, Philips Tech. Rev. 41, 201-223, 1983/84.

^[6] M. Nishiguchi, K. Akagiri and T. Suzuki, A new audio bit rate reduction system for the CD-I format, Proc. 81st Audio Eng. Soc. Conv., Los Angeles, Cal., 1986, reprint No. 2375 (C-4), 11 pp.

and hardware. The software is supplied by companies such as the American firm Microware, who also developed the CD-RTOS control system.

In creating and combining the information for a CD-I, a balance has to be struck between the memory space required on the disc and the quality of the images and sound. High-resolution images require more memory space than low-resolution images. This is also true for sound of CD-DA quality compared with sound that only contains the limited frequency range of speech. Various quality levels for sound and image are therefore defined in the Green Book.

Another important factor is the maximum bit rate available at constant playback speed for the track on the disc — the speed is the same for CD-DA and CD-I and is standardized. At this speed a full-screen video picture of broadcast quality, with sound, can be displayed in less than a second.

Audio

A CD-I player can also be used for playing ordinary Compact Discs with their high-quality audio recordings. The high quality of CD-DA is obtained by sampling the analog audio signal at a sampling rate of 44.1 kHz. The number of bits per sample is 32 for each stereo channel, i.e. 16 for each mono channel.

CD-I has different degrees of compression, as compared with CD-DA, for the digitization of the audio signal. The resulting levels of audio quality are:

- hi-fi quality, with double compression, comparable with the first playing of a conventional long-play disc;
- FM quality, with quadruple compression, comparable with the quality of reception for an FM broadcast signal;
- AM quality, with eightfold compression. This is better than the quality of an AM broadcast signal with no interference.

In general, the compression is obtained not by converting the absolute value of each sample into a binary number, but by converting the difference from the previous sample instead. This is differential pulse-code modulation (DPCM)^[5]. More accurately, a special form of DPCM is used in CD-I; this is adaptive differential pulse-code modulation, or ADPCM^[6]. In ADPCM it is the difference between the actual value

of a sample and a predicted value that is converted into a binary number. At the highest quality level the number is an eight-bit number, at the other levels it contains four bits. The predicted value is obtained from previous samples, with the aid of a prediction function that depends on certain slowly varying characteristics of the signal. Sampling rates of 37.8 kHz for the two highest quality levels and 18.9 kHz for the lowest quality level give the stated values for the bit-stream compression.

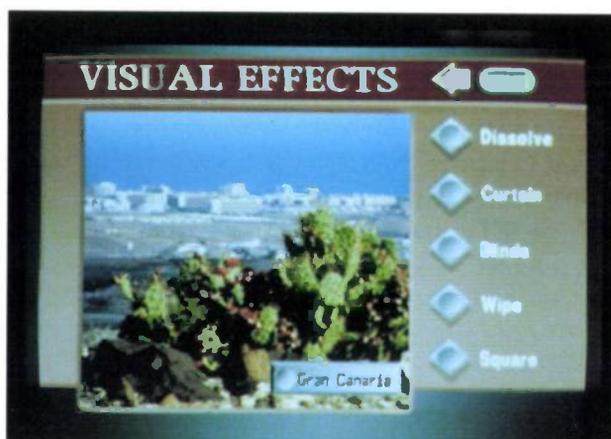
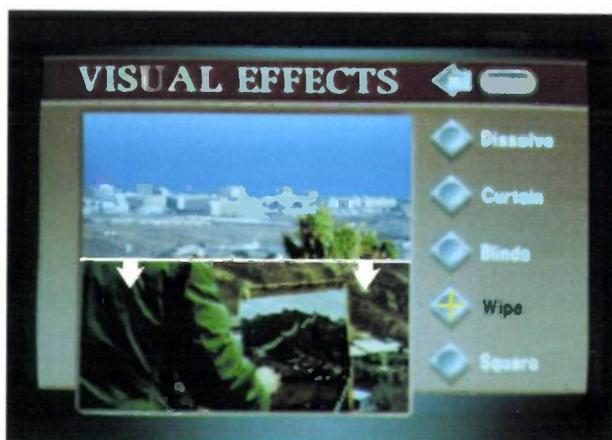
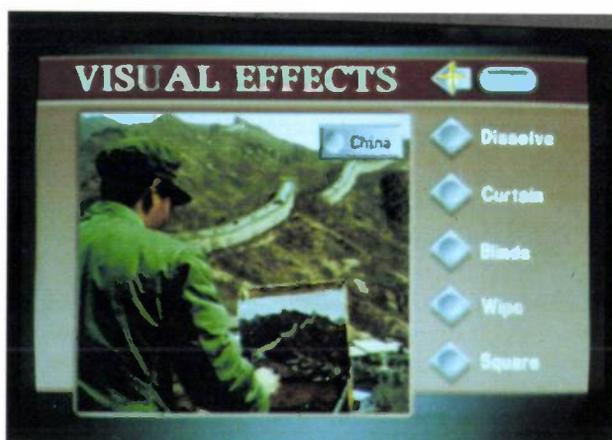


Fig. 5. Combining two image planes (see fig. 4) as a 'Wipe'. The 'China' image is replaced from top to bottom by the 'Grand Canary' image. The menu offers further visual effects with two images:

- 'Dissolve': one image gradually fades into another;
- 'Curtain': one image replaces another from the left and right, like curtains being drawn;
- 'Blinds': one image replaces another as horizontal stripes of increasing width, like a venetian blind;
- 'Square', one image as a square in another image.

A consequence of the bit-stream compression is that as the quality level falls more audio channels become available. For the highest quality level there are 4 channels (i.e. 4 mono channels or 2 stereo channels), for the second level there are 8 channels and for the lowest there are 16 channels. The information is assigned to the channels a sector at a time.

Video

Three different image resolutions are defined in CD-I:

- normal resolution, comparable with the resolution in an ordinary television receiver,
- double resolution, for the presentation of letters and numbers,
- high resolution, in anticipation of future image-display systems or professional applications.

The CD-I player and the MMC module decode the image information read from the disc so that the signal supplied to the monitor or television receiver represents the correct number of lines: 625 at a frame frequency of 25 Hz, or 525 at a frame frequency of 30 Hz.

Since the maximum bit rate is about 1.36 Mbit/s, it will be necessary to wait a few seconds before a picture appears on the screen, unless special precautions are taken with the digitization of the video signal. This is why advanced compression techniques are used. In CD-I there are four methods of image digitization, each appropriate to a particular kind of image material.

- One-dimensional DYUV coding for 'natural' images, such as a colour photograph. In this method the changes in the luminance signal Y and the chrominance signals U and V of successive pixels are converted into binary numbers line by line.
- Direct RGB coding for high-quality graphics images. In this method a five-bit binary number is assigned to each red, green or blue colour component of a pixel. Each colour component therefore has 2^5 intensity values, so that more than 32 000 different colours can be obtained.
- CLUT coding for graphics images that may need to be changed quickly. A Colour Look-Up Table (CLUT) is included on the disc for this application. This table can contain 2^8 , 2^7 , 2^4 or 2^3 different colours. The standard provides a choice from a 'palette' of rather more than 16×10^6 shades. In the equipment now available the choice is limited to about 256 000 shades.
- One-dimensional run-length coding with CLUT, mainly suitable for animation. Here use is made of the knowledge that in this kind of application the colour is usually constant over a large part of a line. The



Fig. 6. Another method of combining image planes; see also fig. 4. Image plane 2 is transparent inside the frame, so that image plane 3, which contains the main menu, becomes visible. Image plane 1 contains the cursor.

binary numbers combine the number of a colour in the table with the number of sequences of pixels in which the colour does not change. Run-length coding can be used to make full-screen moving images.

The CD-RTOS operating system

The information from each sector is also divided up into channels for other kinds of information besides sound. The CD-RTOS operating system ensures that the data stream read from the disc is divided appropriately and sent to different outputs as required. With the information distributed over the channels in this way speech signals in various languages can be included on the disc and therefore in the data stream. When the user chooses a language in his dialogue with the system, CD-RTOS ensures that the appropriate channel is connected to the audio output.

For combining images, the system has four 'image planes'; see fig. 4. Images to be combined are stored temporarily by CD-RTOS in image memories. Various dynamic effects with images can be produced in this way; see fig. 5. Images can also be built up from parts of other images or images can be superimposed. If desired, parts of an image plane can be made 'transparent', so that a lower image plane is made visible; see fig. 6. The lowest image plane can be used as the background.

As stated, each sector can contain audio, video or text information, or software. Information recorded in the subheader of each sector indicates how CD-RTOS should interpret the information in that sector. The address information in the header can be used by the operating system or the user for searching. Interactive searching is a feature of CD-I; see fig. 3.

It will be clear that information in the data stream read from the disc can be interleaved with related information. After the data stream has been sorted out, CD-RTOS sends it to the correct output channel in real time. Facial movements in the image, for example, must correspond exactly with the speech; in other words the information from the video output must be synchronized with that from the audio output.

Current status and further developments

It will have become clear from what we have said that the development of CD-I is a team effort, with contributions from colleagues from various disciplines. The first phase of the development, in which the notable feature was the close cooperation with the Philips research laboratories, included the following activities:

- drawing up a standard,
- producing prototypes,
- specifying integrated circuits in VLSI technology,
- preparation for production.

The activities listed above mainly concern the hardware and the associated software. Experience has shown that it is no use introducing hardware if the data carriers are not obtainable in sufficient variety.

Considerable effort has therefore been put into developing authoring systems and supplying them to companies that make programs for the discs. We want to offer users a wide choice of interesting interactive applications in the near future.

A technical challenge that must soon be faced is that of finding more effective compression techniques, to give further improvement in the quality of moving images. At the same time second-generation integrated circuits will have to be developed. Simpler hardware will then be within reach. The ultimate results will be reductions in price and a corresponding increase in the scale of production, with increasing diversification in hardware and discs.

Summary. The standard for the CD-I system (Compact Disc Interactive) for consumer applications is an extension of the standard for CD-ROM (Read-Only Memory) for professional applications for computers, which in turn is an extension of the standard for CD-DA (Digital Audio). The CD disc contains images, sound, text, and the associated software in digital form. The information is organized in sectors on the disc, each with its own address and a list of contents. There are two levels of error correction, four quality levels for sound and three quality levels for images. This means that quality can be traded against storage capacity and bit rate when the disc is created. The supplier of interactive programs does this by means of an authoring system. The output from the authoring system is the digital information used in manufacturing the 'mother disc'.

1939

THEN AND NOW

1989

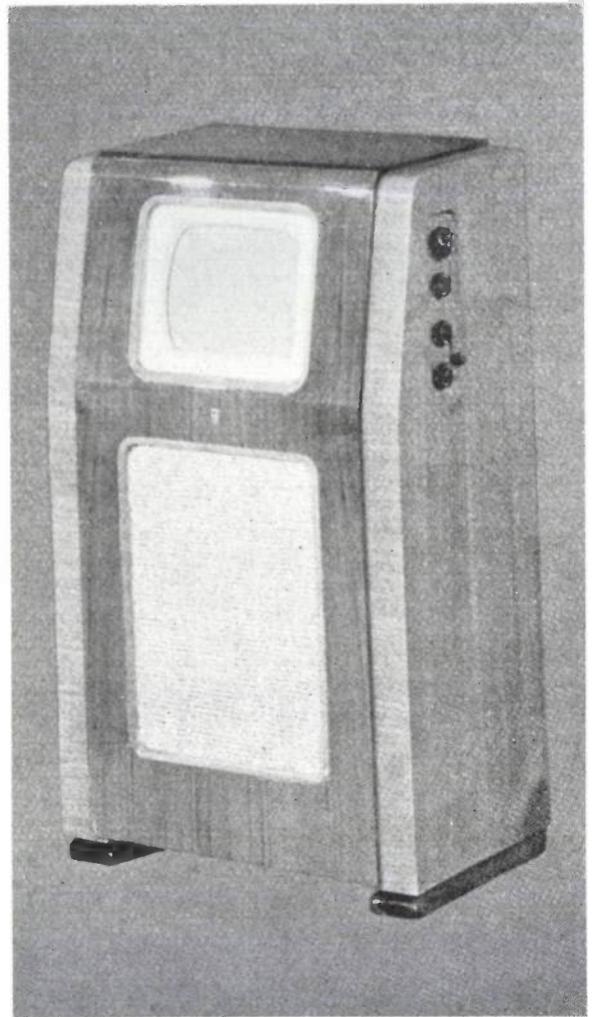
Television receivers

Philips have been designing television receivers for more than fifty years. The console model in the black-and-white photograph ^[*] was 80 cm high, with the cathode-ray tube and the loudspeaker mounted one above the other. The set was tuned to receive the BBC transmissions from London. These provided an interlaced 405-line picture, with 25 pictures a second. In those days the picture tube had rotational symmetry, and the slightly rounded screen face had a diameter of 22 cm. The picture height was 15 cm and the width was 17.5 cm.

Much has changed since that time. The photograph below shows the 28DC 2070 colour television receiver that became available this autumn. The rectangular 625-line picture measures 53 cm by 40 cm; the screen is flat and square and there is not a control to be seen — the set is operated entirely by remote control. The loudspeakers can be positioned separately. Connections are available for video recorder and computer.

The inside has changed too, not just the outside. Most of the discrete components have been replaced by ICs, and the set also contains a number of modules that add new features. One of these is hi-fi stereo sound, made possible by digital signal processing. The receiver also offers PIP (Picture in Picture), which shows a 18 cm by 12 cm picture of another programme simultaneously in a corner of the screen. There is a teletext module, of course, with an 8-page memory in this model.

Work on the television of the future continues, and further changes are just around the corner. One such will be the aspect ratio of the screen — from about 4:3 to 16:9.



The greatest step forward, however, will be the improved viewing with D2-MAC; the resolution will be better, and there will be fewer artefacts, since there will be less crosstalk between the chrominance and luminance signals in the transmitted signal. And the HD-MAC standard will be introduced later, with a 1250-line picture.

[*] From Philips Technical Review, December 1939.

Phosphor screens in cathode-ray tubes for projection television

R. Raue, A. T. Vink and T. Welker

Cathode-ray tubes with phosphor screens have been used for picture display since the early days of television. The conventional direct-view tube for colour television contains a screen with three phosphors that give red, green and blue light when they are excited by electrons. In projection television, colour display is obtained by superimposing the magnified images from three separate tubes on a large viewing screen. The requirements for the phosphor screens are much harder to meet than in a direct-view tube, because of the high excitation densities. The renewed interest in projection television will certainly be strengthened by recent improvements in phosphor screens to give much better picture quality.

Introduction

The phosphor screens in the cathode-ray tubes in television receivers convert the energy of fast electrons into light. In conventional direct-view television the viewer looks directly at the picture produced on the tube faceplate, but in projection television the image is projected on to a large viewing screen — an attractive way of displaying high-quality pictures with a diagonal of 100 cm or more. For projection television in colour, three separate cathode-ray tubes produce images in the primary colours red, green and blue. These images are projected in exact register with the aid of electronic control.

Recent improvements in picture quality have led to an increased interest in projection television, and this will receive a further stimulus from the introduction of high-definition television (HDTV) with large picture formats. In a previous article in this journal^[1] it was shown how the brightness and colour rendering (chromaticity) can be improved by applying interference filters to curved faceplates. In this article we shall discuss the phosphor screens to be used.

A phosphor screen in a projection-television tube is used in much the same way as the screen in a direct-view tube. A phosphor layer deposited on the face-

plate is covered by a thin aluminium film that acts as an electrode and also reflects light forwards. Electrons arriving from the back of the screen pass through the aluminium film and into the phosphor layer. The light generated by the electron excitation leaves the phosphor screen from the front after reflection and scattering.

Pictures with sufficient brightness and resolution for projection television are obtained by excitation with an intense and very small electron spot. The diameter of this spot is about a tenth of the diameter of the spot in present direct-view tubes, and the maximum excitation density ($\approx 2 \text{ W/cm}^2$) is about a hundred times higher. This sets very difficult requirements for the phosphor screens. If the screen is to give optimum performance, it is necessary to consider the luminescence properties of phosphors under the conditions encountered in projection television, the suitability of a phosphor for screen preparation, and the optical properties of the screen.

Properties of importance in projection television are the energy-conversion efficiency (particularly at high excitation densities), the 'chromaticity coordinates', the decay time and the thermal quenching,

Dr R. Raue and Dr T. Welker are with Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany, and Dr Ir A. T. Vink is with Philips Research Laboratories, Eindhoven.

^[1] L. Vriens, J. A. Clarke and J. H. M. Spruit, Interference filters in projection television tubes, Philips Tech. Rev. 44, 201-210, 1989.

and the variation in properties under prolonged intense electron bombardment. The requirements for chromaticities and decay time are the same as in direct-view tubes. This means that the chromaticity coordinates of the phosphors (including the effect of interference filters) should meet the European Broadcasting Union specifications for red, green and blue to ensure faithful colour reproduction, and that the decay time should be short enough to avoid 'smearing' in moving images. The efficiency should be as high as possible. Some of the efficient phosphors used in direct-view television are not suitable for projection television, however, because the light output saturates strongly at high excitation densities. It is therefore also necessary to consider phosphors that are less efficient at low excitation densities but better at high densities because they are more linear. The requirements relating to thermal quenching and deterioration in efficiency are also more difficult than in direct-view tubes.

The suitability of a phosphor for use in a screen depends on the type of layer required. A typical screen has an average thickness between 10 and 30 μm and consists of a powder with a mean grain size between a few μm and about 10 μm . It is also necessary to optimize the size distribution about the mean grain size. The morphology of the grains should not be so complex that it is impossible to obtain the high packing density necessary for stability at high excitation densities and for good optical properties of the screen. Finally, the phosphor must not be affected by the screen processing.

The optical properties of a screen depend closely on the thickness of the phosphor layer. At a given packing density of the phosphor, the layer must be thick enough to ensure sufficient absorption of the electron energy. A limit to the thickness is set, however, by multiple light scattering in the layer. In a thick layer much of the light is lost owing to scattering and subsequent absorption in the phosphor screen, and the sharpness of the image (resolution) is reduced by lateral scattering. However, some scattering is necessary to give a gain in brightness in the forward direction. This scattering depends on the size and morphology of the grains and on the index of refraction of the phosphor. In current phosphors the grain size is the dominant parameter. Prolonged electron bombardment often leads to additional light absorption due to 'browning' of the screen; its effect is increased by multiple light scattering. The browning may be affected by the screen processing.

In this article we shall compare phosphors for projection television with phosphors used for direct-view screens, and we shall present a model for the analysis

and prediction of optical screen properties. We shall also review the screen processing and discuss the degradation of phosphor screens in projection-television tubes. Finally, we shall look briefly at the future prospects.

Phosphors for projection television

The physical processes involved in the electron excitation of a phosphor are presented schematically in *fig. 1*. A phosphor basically consists of a 'host lattice' and an 'activator', a general term for elements incorporated to activate the luminescence. Fast electrons penetrate the phosphor and are slowed down by the increasing interaction with the host lattice. The penetration depth increases with the acceleration voltage and decreases with the density of the phosphor; typical values of the penetration depth for 30-kV electrons, as used in projection television, range from 3 to 5 μm . The electrons excite inner shells of the host-lattice atoms and the resulting energetic electrons generate secondary electrons. Eventually electron-hole pairs are formed in the host lattice and these can transfer their energy to the activator ions, which are then excited to emit light. Unfortunately there are also a number of loss processes such as energy transfer to 'killer centres', surface recombination at the 'dead layer', and non-radiative deactivation in the activator ions.

Ideally the three phosphors that emit the red, green or blue light should have high efficiencies up to high

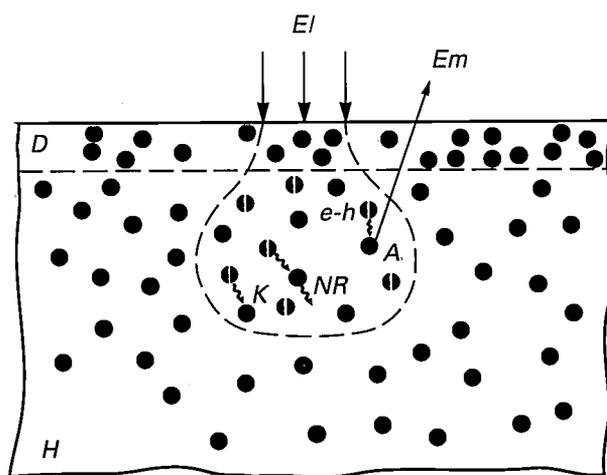


Fig. 1. Diagrammatic representation of the luminescence produced in a phosphor grain by electron excitation. Accelerated electrons EI are incident on a host lattice H containing activator ions A . The incident electrons are absorbed in the host lattice; the penetration range increases with the acceleration voltage. The absorption of electron energy results in the generation of electron-hole pairs $e-h$. These can transfer their energy to the activator ions, which can be excited to give light emission Em . No luminescence is generated in a thin region near the surface, the 'dead layer' D . Unwanted processes also arise, such as energy transfer to killer centres K and non-radiative deactivation NR in the activator ions.

excitation densities, as well as the correct chromaticities and decay times. Thermal quenching and deterioration with time should also be negligible. In practice the choice is a compromise, mainly determined by the balance between the efficiency and linearity of the phosphor, provided the other properties are acceptable. It can therefore happen that the 'best choice' may change with time because of changing tube and system requirements in the future, e.g. for the scanning conditions. It is therefore important to present the luminescence properties in such a way that the performance of the tubes for blue, green and red in various combinations can be predicted. We shall now go on to discuss the basic essentials of the spectral properties and efficiency of the luminescence, sub-linearity at high excitation densities, and an experimental comparison of some of the phosphors that could be used in projection television.

Spectral properties and efficiency

The combination of host lattice and activator determines the luminescence properties of a phosphor. Emission spectra for several phosphors are shown in *fig. 2*. The blue emission of ZnS:Ag (also used in direct-view tubes) is broad; this is ascribed to the strong interaction between the silver activator and the host lattice. In the same way the green-emitting direct-view phosphor $\text{Zn}_{0.95}\text{Cd}_{0.05}\text{S}:\text{Cu}$, indicated by (Zn,Cd)S:Cu, gives a broad-band emission. For projection television, however, a green-emitting Tb^{3+} -activated oxidic phosphor is preferred. This phosphor emits in spectral lines, as does the phosphor $\text{Y}_2\text{O}_3:\text{Eu}$, used for red in projection television.

The line emission of phosphors activated with Tb^{3+} or Eu^{3+} indicates that the interactions between these activators and the host lattice are very weak [2]. These interactions are very weak, because of the special nature of these rare-earth ions: the luminescence transitions take place inside their inner 4f shells, which are well shielded from the environment by outer electron shells. The mutual interactions between these ions are also very weak, which means that they can be used at concentrations of up to 1-10% without serious loss in efficiency due to concentration quenching. Their behaviour is different from ZnS:Ag and (Zn,Cd)S:Cu, where concentration quenching limits the useful activator concentration to about 0.05%.

In our discussion of the energy-conversion efficiency we shall first consider the various stages in the excitation process; see *fig. 3*. Electron-hole pairs are generated by the incident fast electrons [3], and it has been found that the generation of one thermalized electron-hole pair requires an average energy βE_g ,

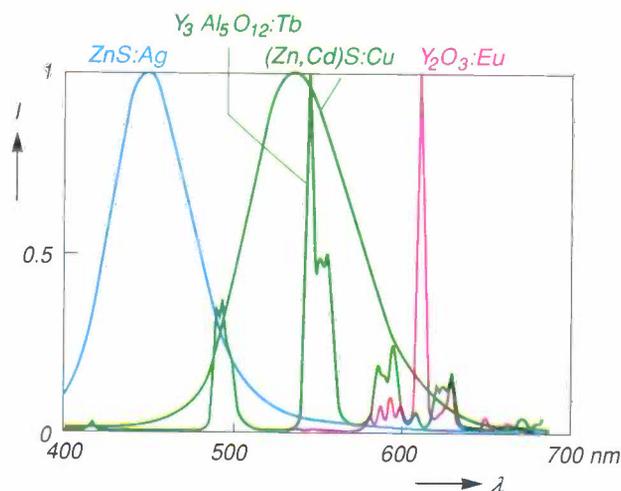


Fig. 2. Emission spectra produced in phosphors by cathode-ray excitation. The normalized photon intensity I is shown as function of the wavelength λ . The spectra of the blue-emitting ZnS:Ag and the green-emitting (Zn,Cd)S:Cu are broad; the spectra of the green-emitting $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Tb}$ and the red-emitting $\text{Y}_2\text{O}_3:\text{Eu}$ consist of sharp lines.

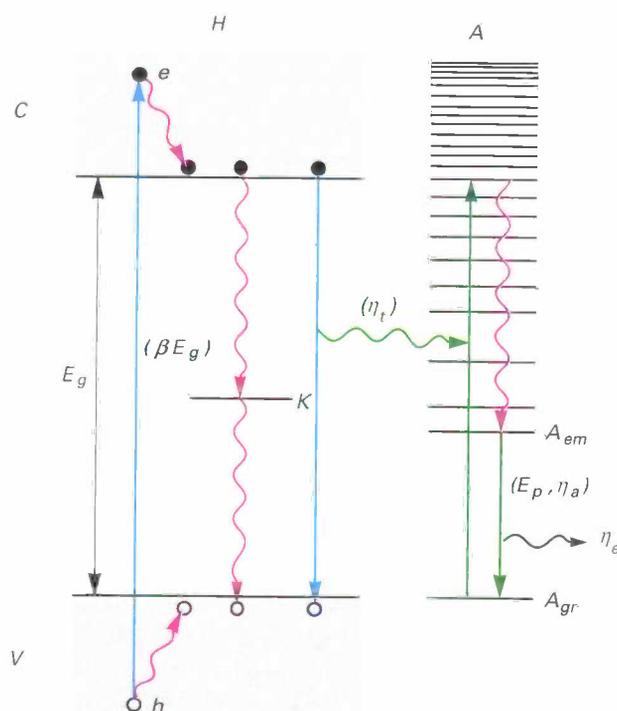


Fig. 3. Different stages in the electron excitation of a host lattice H with activator ions A . The generation of a thermalized electron-hole pair formed by a hole h in the valence band V and an electron e in the conduction band C requires a mean energy βE_g , where E_g is the band gap of the host lattice. Electrons in the conduction band can recombine with holes in the valence band via killer centres K or via energy transfer (at an efficiency η_i) to the activator. The activator returns from its excited state to the emitting state A_{em} via non-radiative transitions. The return to the ground state A_{gr} then causes the emission of light at a photon energy E_p , a quantum efficiency η_a (the ratio of the number of photons generated to the number of activator ions excited) and a photon-escape efficiency η_e .

[2] See for example G. Blasse and A. Brill, Characteristic luminescence, Philips Tech. Rev. 31, 303-332, 1970.

[3] D. J. Robbins, On predicting the maximum efficiency of phosphor systems excited by ionizing radiation, J. Electrochem. Soc. 127, 2694-2702, 1980.

where E_g is the band gap and β ranges from about 2.7 to 5, depending on the host lattice. The next step is the transfer of the electron-hole energy to the activator, with an efficiency of η_t . The excited activator then relaxes to its luminescing state and returns to its ground state. This leads to the emission of a photon of energy E_p , with a quantum efficiency η_a and an efficiency η_e for escape from the phosphor powder or phosphor screen.

If we consider the different stages we can estimate the energy-conversion efficiency η_{CR} for cathode-ray excitation:

$$\eta_{CR} = \frac{E_p}{\beta E_g} \eta_t \eta_a \eta_e \quad (1)$$

The maximum efficiency is obtained when η_t , η_a and η_e are all unity. It is advantageous if the values of β and E_g are as small as possible, but the value of E_p is fixed by the emission colour required. For ZnS:Ag we have $E_g = 3.8$ eV, $E_p = 2.7$ eV and $\beta = 2.7$, so that the theoretical maximum of η_{CR} is 26%. The highest efficiencies obtained with this phosphor are close to this value. Oxidic host lattices such as $Y_3Al_5O_{12}$ have larger values for E_g and β , and therefore have lower efficiencies. The highest value obtained with green-emitting $Y_3Al_5O_{12}:Tb$ is about 10%, whereas with (Zn,Cd)S:Cu efficiencies of up to 18% can be obtained. These efficiency values are associated with the conversion of the absorbed electrons; the conversion efficiency of the incident electrons is only 80 to 90% of these values because of electron back-scattering.

Since the excitation densities are high in projection television, the screen temperature can rise significantly, and may even reach 100 °C. This implies that the decrease in efficiency with increasing temperature (thermal quenching) should be small. This quenching is mostly due to non-radiative losses from the emitting state and depends on the coupling to the lattice vibrations [2]. Although models are available that explain these processes, it is difficult to make accurate predictions. However, a knowledge of the measured temperature dependence of the efficiencies enables us to select phosphors with negligible losses at 100 °C. A more serious problem with high excitation densities is the sublinearity of the luminescence output.

Sublinearity

To demonstrate the sublinearity we show the efficiency of several phosphors in *fig. 4* as a function of the excitation density, i.e. the energy density of the excitation pulse, at typical values for projection tubes and direct-view tubes. The efficiencies of phosphors such as ZnS:Ag and (Zn,Cd)S:Cu fall off considerably at higher excitation densities. On the other hand

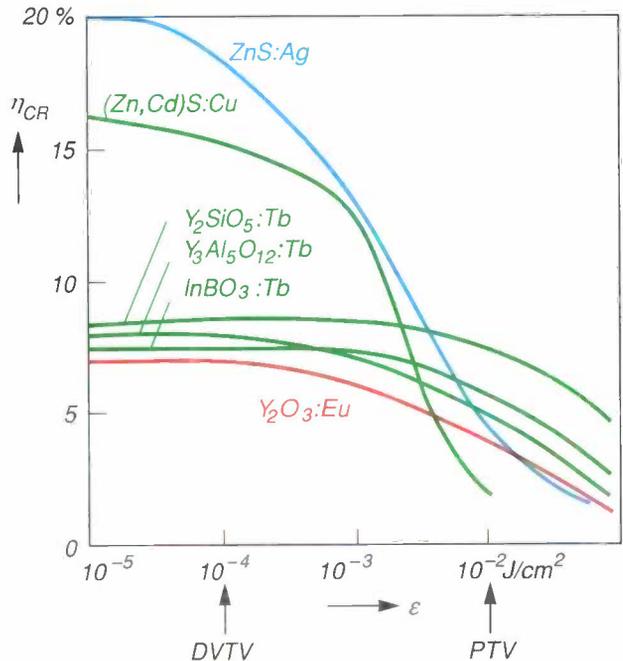


Fig. 4. Energy-conversion efficiency η_{CR} of various phosphors as a function of the excitation loading, i.e. the excitation density per pulse (ϵ), in a projection-television tube at a fixed electron spot diameter. The efficiency of the phosphors ZnS:Ag and (Zn,Cd)S:Cu, which is high at low excitation loadings, decreases strongly at higher loadings. The arrows indicate typical values of ϵ in direct-view television tubes (DVTV) and in recent projection television tubes (PTV).

the Tb^{3+} and Eu^{3+} phosphors have lower efficiencies at lower densities, but at higher densities they have the advantage of superior linearity.

As long ago as 1949 [4] a model was proposed that would explain the sublinearity of phosphors such as ZnS:Ag. We can describe the basic features of this model with the aid of *fig. 3* and *eq. (1)*, and note that the electron beam scans the phosphor screen in lines, thus giving excitation of each picture element in short pulses. The pulses have a typical duration of 200 ns, and are repeated at 40-ms intervals. Most phosphors have a decay time that is much longer than the pulse duration, but much shorter than the pulse interval. This means that activator ions excited by a pulse all return to the ground state before the next pulse arrives. As the excitation density increases, the number of excited activator ions increases until all the activator ions are excited. The maximum averaged photon flux per unit phosphor volume is equal to the ratio of the activator concentration to the pulse interval. A high activator concentration is therefore required for linearity at high excitation densities. This largely explains why Tb^{3+} and Eu^{3+} phosphors with an activator concentration of 1-10% are far more linear than ZnS:Ag with an activator concentration of only 0.05% or less.

Further investigations showed, however, that this model is too simple. It does not explain the differences in linearity between various Tb^{3+} phosphors, for example. A more detailed analysis reveals that interactions between close activator ions in excited states initiate additional loss processes at high densities. The most direct proof of this is given by the luminescence decay [5], as shown in *fig. 5*. At low densities only the radiative decay of the activator is observed, but high densities give an initial shorter decay from which the additional loss can be derived. After some time the close activators that have been excited will have decayed, so that the slope again corresponds to the radiative decay.

The effect of interactions in excited states cannot be predicted very accurately, but must be determined experimentally. A general problem with measurements on projection-television tubes is the variation in the size of the electron spot with current. It is difficult to measure the spot size because of effects due to light scattering and sublinearity. A better control of the conditions can be obtained by measuring the luminescence output for short excitation pulses with a well-defined spot profile; we used a scanning electron microscope. The optical output can now be determined accurately as a function of the excitation density per pulse, given by $Vj\Delta t$, where V is the acceleration voltage, j the current density and Δt the pulse duration. At a fixed voltage, a curve of the output per pulse against $j\Delta t$ always has the same shape for a given phosphor, provided that Δt is much less than the initial decay time [6]. Although the values of j and Δt differ from the values in actual tubes, curves like the ones shown in *fig. 4* can be used for comparing phosphors and predicting tube performance [7].

Comparisons of phosphors

In recent years many phosphors have been investigated in our laboratories to see if they would be suitable for application in projection-television tubes. The investigations combined the evaluation of well-known phosphors with a search for new phosphors. Data relating to the properties mentioned above will now be quoted for a number of phosphors.

Table I gives values for the energy-conversion efficiency and luminous efficacy at low excitation densities, the thermal quenching at 100 °C, the decay time and the sublinearity. The luminous efficacy is obtained by multiplying the energy-conversion efficiency by the lumen equivalent, i.e. the ratio of the number of lumens to the emitted power, calculated from the measured spectral distribution. The sublinearity can be conveniently characterized by two parameters: the relative efficiency at the average energy density per

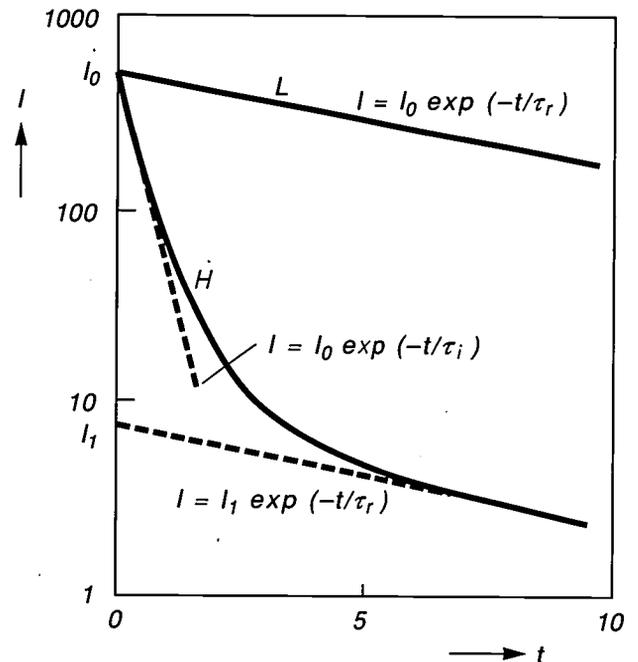


Fig. 5. Effect of the excitation density on the luminescence decay of a phosphor. The luminescence intensity I , on a logarithmic scale, is plotted against the time t , both in arbitrary units. The intensities at the end of the excitation pulses at $t=0$ are set equal to I_0 . A low-density pulse gives curve L : a straight line of slope $1/\tau_r$, where τ_r is the radiative decay time of the phosphor. A high-density pulse gives curve H , with an initial shorter decay characterized by a slope $1/\tau_i$ followed by a tail with a slope $1/\tau_r$.

pulse, typically 10 mJ/cm² in present tubes, and the density at which the efficiency is half the efficiency in the linear regime.

The only phosphor we consider for blue is ZnS:Ag. Even though it is very sublinear it is still the only phosphor suitable for projection television, since there is no proven alternative. Its sublinearity is a great disadvantage in blending red, green and blue to produce white light, and the projection system has to be adapted to deal with this problem.

The green Tb^{3+} phosphors are much more linear than (Zn,Cd)S:Cu, and $Y_2SiO_5:Tb$ and $Y_3(Ga,Al)_5O_{12}:Tb$ are among the best. The choice of the Tb^{3+} phosphor also depends on the decay time, the available screen technology and the degradation with time.

[4] A. Bril, On the saturation of fluorescence with cathode-ray excitation, *Physica* **15**, 361-379, 1949;

A. Bril and F. A. Kröger, Saturation of fluorescence in television tubes, *Philips Tech. Rev.* **12**, 120-128, 1950.

[5] D. M. de Leeuw and G. W. 't Hooft, Method for the analysis of saturation effects of cathodoluminescence in phosphors; applied to $Zn_2SiO_4:Mn$ and $Y_3Al_5O_{12}:Tb$, *J. Lumin.* **28**, 275-300, 1983.

[6] D. B. M. Klaassen, T. G. M. van Rijn and A. T. Vink, A universal description of the luminescence saturation behaviour per phosphor, *J. Electrochem. Soc.*, **136**, 2732-2736, 1989.

[7] T. Doyle, D. B. M. Klaassen and M. J. G. Lammers, The influence of high scanning frequencies on the luminescence saturation properties of phosphors for CRT projection systems, *IEEE Trans.* **ED-36**, 1876-1881, 1989.

Table I. Comparison of blue-, green- and red-emitting phosphors for projection television. The Table shows the energy-conversion efficiency η_{CR} and luminous efficacy η_L for cathode-ray excitation at room temperature and low excitation densities, the ratio r_{th} of the efficiency at 100 °C to the efficiency at room temperature, the luminescence decay time τ , and the parameters characterizing the linearity at high excitation densities, i.e. the ratio r_{ex} of the efficiency at 10 mJ/cm² to the efficiency at low excitation densities, and $\varepsilon_{1/2}$, the excitation density at which the efficiency has been halved.

Phosphor	$\eta_{CR}^{[a]}$ (%)	η_L (lm/W)	r_{th}	τ (ms)	r_{ex}	$\varepsilon_{1/2}$ (mJ/cm ²)
<i>Blue</i>						
ZnS:Ag	20	13	0.89	0.01-0.07 ^[b]	0.23	2
<i>Green</i>						
(Zn,Cd)S:Cu	16	85	1.00	0.01-0.05 ^[b]	0.12	2
Y ₃ Al ₅ O ₁₂ :Tb	8	35	1.00	3	0.61	20
Y ₃ (Al,Ga) ₅ O ₁₂ :Tb	9	42	0.97	3	0.82	70
Y ₂ SiO ₅ :Tb	9	41	0.94	2	0.88	95
LaOCl:Tb	10	45	0.98	1	0.70	25
InBO ₃ :Tb	8	42	1.00	7.5	0.76	45
Gd ₂ O ₂ S:Tb	11	48	0.80	0.7	0.51	10
<i>Red</i>						
Y ₂ O ₂ S:Eu	13	25	0.68	0.5	0.40	8
Y ₂ O ₃ :Eu	7	22	0.91	2	0.59	12

^[a] Ratio of the luminescence energy output to the electron-energy input, of interest for practical tubes. If just the absorbed energy is considered higher values are obtained, since 10-20% of the incident electron energy is not absorbed because of back-scattering.

^[b] Decreases with increasing excitation density.

Red phosphors include Y₂O₂S:Eu, used in direct-view tubes, and Y₂O₃:Eu. Although Y₂O₂S:Eu has a higher luminous efficacy at low densities, superior thermal properties and linearity make Y₂O₃:Eu the obvious choice for projection-television tubes.

The efficiencies and spectral data in the above comparisons represent intrinsic phosphor properties. However, to predict and assess the tube performance we also need to know the optical properties of phosphor layers in screens.

Optical properties of phosphor screens

Thin-film screens and powder screens have been considered for projection television. Thin-film phosphor screens are prepared directly on the substrate by deposition techniques such as epitaxy and evaporation. The screens are transparent and give excellent resolution and stability at high excitation densities. A major disadvantage, however, is their low photon escape in the forward direction due to internal light trapping and reflection and refraction losses at the interface with a medium of lower refractive index. Expensive substrates are also necessary: these have to be of single-crystal material for epitaxy, or unaffected by the high-temperature annealing required for va-

pour-deposited screens, for example. Thin-film screens have therefore been used mainly for professional applications where high resolution and contrast are required and the low efficiency can be compensated by high excitation densities.

Powder screens are preferred for consumer applications. These screens have to be prepared in two separate steps: the phosphor is prepared first and then deposited on the substrate. All thermal treatments at temperatures above 450 °C are carried out in the first step, so that less expensive substrates can be used. A characteristic feature of powder screens is multiple light scattering, which counteracts the decrease mentioned earlier in the photon escape and gives a forward gain in intensity instead. From now on we shall only consider powder screens.

The optical properties of a phosphor screen (*fig. 6*) are basically determined by the scattering and absorption in the powder layer and the reflectance of the aluminium film. It is not easy to formulate a theory for the light propagation, because of the complex topography and multiple scattering. Successful phenomenological calculations^[8] have been restricted to one dimension and only give the output integrated over angle and position. We have developed a computer model for light propagation in three dimen-

sions^[9]. The model gives complete information about the angular and spatial distribution of the light output and can be used for modelling the optical properties.

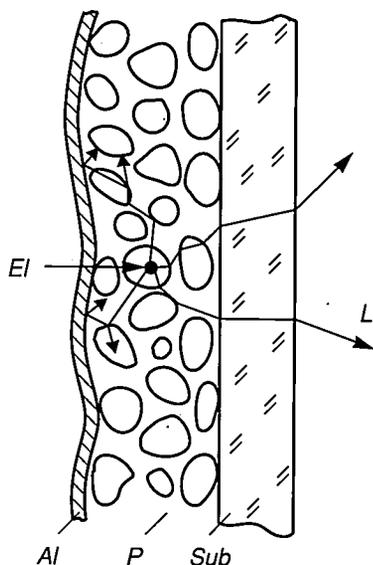


Fig. 6. Diagram of the cross-section of a phosphor screen for projection television. The phosphor is deposited as a powder layer *P* on a transparent substrate, typically a glass window a few mm thick. The layer is covered by a thin aluminium film *Al* (thickness $0.2 \mu\text{m}$). Electrons *El* enter the phosphor layer from the aluminium side and the light generated *L* leaves the screen on the substrate side.

The model

The computer model is based on the continuum approximation using statistically distributed scattering and absorption centres with variable mean densities. Other input parameters are the layer thickness and the reflectance of the aluminium film. The effect of the substrate is neglected here, but the model has since been extended to include substrate effects such as the loss of contrast due to lateral light propagation ('halo effect') and the gain in forward intensity with interference filters^[1].

In the simulations it is assumed that electrons with a typical penetration depth of $3 \mu\text{m}$ generate photons with an isotropic angular distribution. The propagation of the light is simulated by the Monte Carlo method: paths are traced for individual photons in the presence of the scattering and absorption centres. The probability of scattering is expressed by the mean pathlength between two scattering events; the angular probability is assumed to be isotropic. The probability of absorption is expressed by the mean pathlength of a photon before it is absorbed. Typical absorption lengths are larger than the scattering length by three orders of magnitude. Most photons therefore escape from the screen after several scattering events and

reflections at the aluminium film. Although absorption in the phosphor only plays a minor part, it may become important after prolonged electron bombardment. Photons escaping from the screen are classified by position and angle of emergence.

Results

The emission intensity from a perfectly diffusing surface has a 'Lambert distribution': the intensity is proportional to the cosine of the angle θ between the normal to the surface and the direction of observation. The angular distribution from a phosphor powder screen, however, is affected by multiple scattering. Fig. 7 shows the effect of different values of the scat-

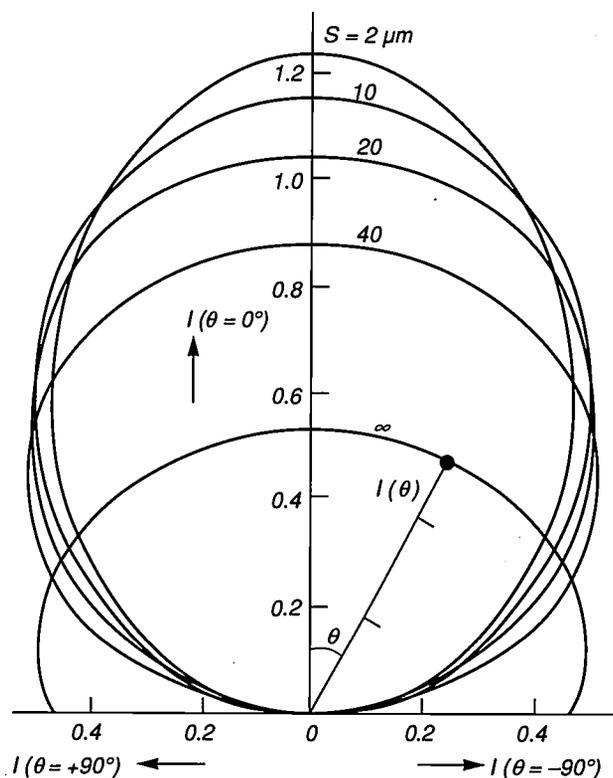


Fig. 7. Angular intensity distribution of a phosphor powder screen for cathode-ray excitation at five different values of the scattering length *S* and fixed values of the layer thickness ($20 \mu\text{m}$), absorption length (10 mm) and aluminium reflectance (90%). The intensity *I* is shown as a function of the angle θ to the normal, and its scale is normalized with respect to the Lambert distribution (see text). If the scattering is negligible (i.e. *S* is very large) the angular distribution is almost isotropic, with the effect of absorption visible at large angles. An increase in the scattering (i.e. a decrease in *S*) gives an increase in gain in the direction of the normal. At $S < 20 \mu\text{m}$, this gain is higher (up to 22%) than expected from the Lambert distribution.

[8] A. Brill and H. A. Klasens, Intrinsic efficiencies of phosphors under cathode-ray excitation, Philips Res. Rep. 7, 401-420, 1952.

[9] W. Busselt and R. Raue, Optimizing the optical properties of TV phosphor screens, J. Electrochem. Soc. 135, 764-771, 1988.

tering length S , with the intensity scale normalized to the Lambert distribution. At $S = 20 \mu\text{m}$ the intensity in the forward direction ($\theta = 0^\circ$) corresponds to the Lambert distribution. With increased scattering (smaller values of S) the intensity in the forward direction is larger: up to 22% at $S = 2 \mu\text{m}$. If the scattering is negligible (very large S) the angular distribution is almost isotropic.

The optical performance of the phosphor screen depends on the amount of light emitted into the aperture of the projection optics. Increasing the scattering introduces two opposing effects, as shown in *fig. 8* for light emission within an acceptance angle of 50° . The forward gain in intensity gives an increase, whereas the angle-integrated output decreases because of the increased number of reflections at the aluminium film, which in turn leads to increased absorption losses and therefore to a lower photon-escape efficiency of the screen. As a result, the light output within the acceptance angle has a maximum at a scattering length of about half the layer thickness. Calculations with different screen parameters have shown that the optimum layer thickness is always about twice the scattering length. It was possible to show that this is the case if the layer thickness corresponds to about 2.5 times the mean grain size^[9]. We should note, however, that this result only takes the optical properties into account. For optimum brightness the electron energy must be absorbed sufficiently. This means that there should be a minimum phosphor coverage for a given acceleration voltage.

As mentioned earlier, light scattering also affects the resolution. One measure of the loss of resolution is the line-spread function, i.e. the screen response to an infinitesimally narrow line. The calculated linewidth of the line-spread function at 5% of peak height increases linearly with layer thickness: see *fig. 9*. Calculations with varying scattering lengths, absorption lengths and aluminium reflectances have shown that these parameters have little effect. This has also been found experimentally for a wide range of screens with different phosphors, grain-size distributions and morphologies. Measured linewidths at 5% of peak height are also shown in *fig. 9*. To give a comparison with the calculated results we have also plotted the theoretical linewidth obtained when the finite resolution of the experimental arrangement is taken into account. The agreement between experimental and calculated linewidths confirms that the layer thickness is the dominant parameter; screens for high resolution should be as thin as possible. Screens with a sublinear phosphor such as ZnS:Ag have an additional loss of resolution at high excitation densities because of the associated increase in optical spot size^[7].

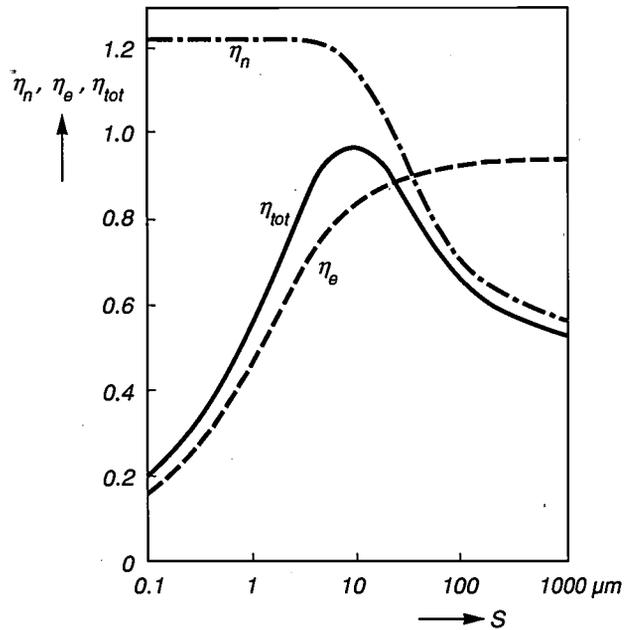


Fig. 8. Increasing the scattering (decreasing S) in a phosphor screen introduces two opposing effects on the amount of light emitted within a given acceptance angle (50°): the intensity with respect to the Lambert distribution (η_n) increases from about 0.5 to 1.22, whereas the angle-integrated screen efficiency (the photon-escape efficiency η_e of the screen) decreases because of the higher absorption losses. The resulting total screen efficiency η_{tot} has a maximum at $S \approx 10 \mu\text{m}$.

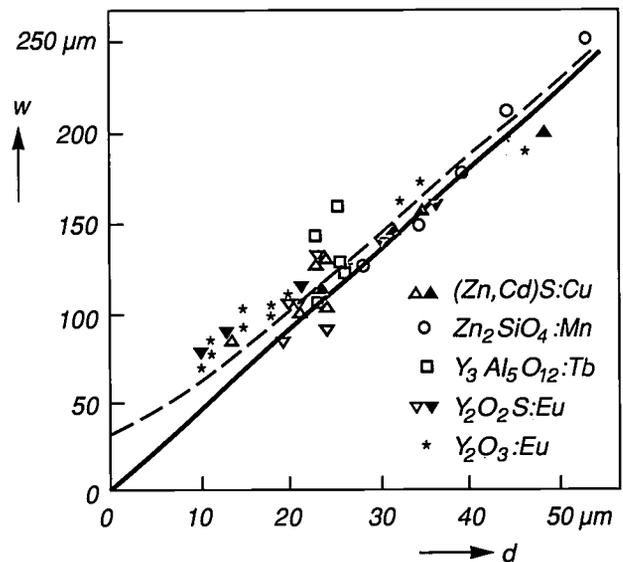


Fig. 9. Linewidth w of the line-spread function at 5% of the peak height, as a function of the layer thickness d of powder screens of several green-emitting and red-emitting phosphors with cathode-ray excitation. Some of the open circles represent results for different batches of the same powder varying in grain size; the lower linewidths correspond to smaller grains for a given layer thickness. Calculated curves are also shown (see text); in the upper curve the broadening due to the experimental arrangement has been taken into account. The agreement between experiment and theory is good; this proves that the linewidth is mainly determined by the layer thickness.

Screen processing

In screen processing it is important to be able to produce a thin layer. For an electron penetration depth of $3\ \mu\text{m}$ and a maximum phosphor packing density of 60%, the minimum thickness should be about $5\ \mu\text{m}$. Since the average thickness of an optimum screen is about 2.5 grains, the minimum grain size should be about $2\ \mu\text{m}$. However, it is not easy to prepare such small grains with a high luminescence efficiency and a good yield, and they are difficult to handle in screen processing. This means that larger grains are often preferred, but here again there is a maximum, determined by the loss of resolution with increasing grain size and screen thickness. In the tubes now manufactured for the consumer market the mean grain sizes range up to about $10\ \mu\text{m}$. The size distribution should be narrowly centred around the mean grain size for a homogeneous screen.

The morphology of the grains is also important. Since the phosphor must be stable at high excitation densities, small single crystals are preferred. They should be simple in shape to permit close packing. The grains must also be chemically stable so that they are not affected by the screen processing. This is usually carried out in an aqueous basic or acid environment, and requires heat treatments at up to $450\ ^\circ\text{C}$.

In screen preparation a phosphor layer and a high-reflectance aluminium film must be deposited. The aluminium film is usually vapour-deposited on a polymer film previously applied to the phosphor layer. This prevents the aluminium from penetrating the phosphor, which would cause severe optical absorption. After the aluminium deposition the polymer film is removed by baking the screen.

Both wet and dry processes can be used for depositing the phosphor layer. In dry processes the substrate is first coated with adhesive, and the phosphor powder is dusted on to the adhesive coating. The powder may be mechanically mixed with the adhesive to improve layer homogeneity and adhesion. Photosensitive adhesives can also be used. With an appropriate binder, the powder can be processed to form a paste that can be printed on to the substrate directly (screen printing). In wet processes a phosphor suspension with a dispersing agent is prepared. This suspension can be added to a bath with the substrate at the bottom; the phosphor grains sink to the substrate where the powder layer is formed (sedimentation). If the grains are charged, the deposition can be induced by an electric field (electrophoresis). Suspension can also be spread over the substrate by tilting and rotating (flow coating). In all processes a binding agent must be used, either in the bath or on the substrate.

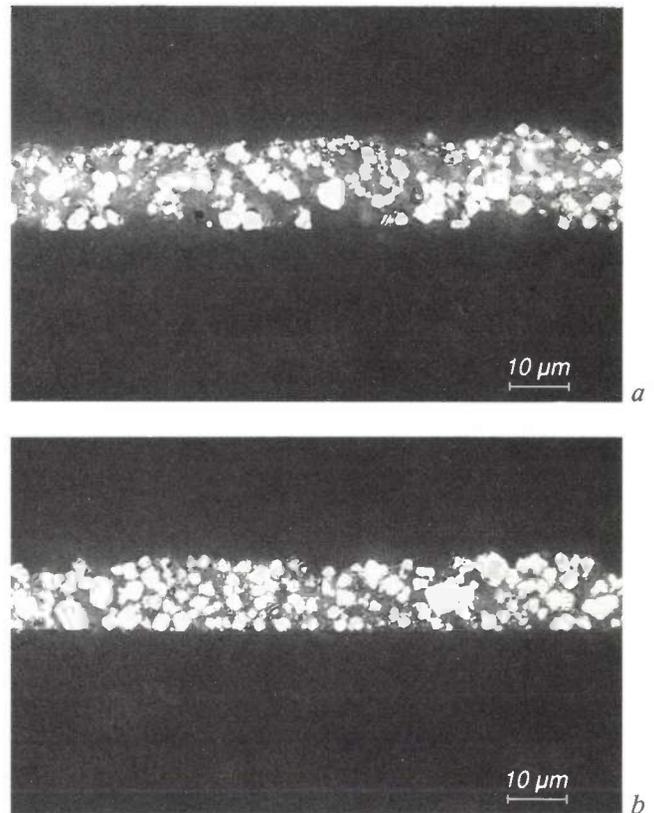


Fig. 10. Cross-sectional view of powder screens containing $\text{Y}_2\text{O}_3:\text{Eu}$, prepared by a standard process (a) and by an improved process (b) giving a much higher packing density.

The most important process for the preparation of screens for projection television is sedimentation. This is a simple technique and gives screens with good brightness, resolution and operating life. The renewed interest in screens for high-definition television (HDTV) has led to some modifications in the process to give improved resolution and life but without loss of brightness.

The sedimentation process is characterized by geometrical parameters such as the height of the liquid in the sedimentation bath, and also by the binder chemistry, which has a considerable influence on the final screen performance and the chemical stability under electron excitation. In a preferred process the phosphor grains are dispersed in a silicate solution that forms part of the binder system and also gives the grains a negative charge and prevents them from sticking together ('agglomeration'). The suspension is poured into the sedimentation vessel with an aqueous electrolyte to act as the coagulant. The positive ions in the electrolyte induce binding between the charged grains and between the grains and the substrate.

The screen structure can be affected by the composition and pH-value of the dispersion solution and the sedimentation bath. Fig. 10 shows cross-sections

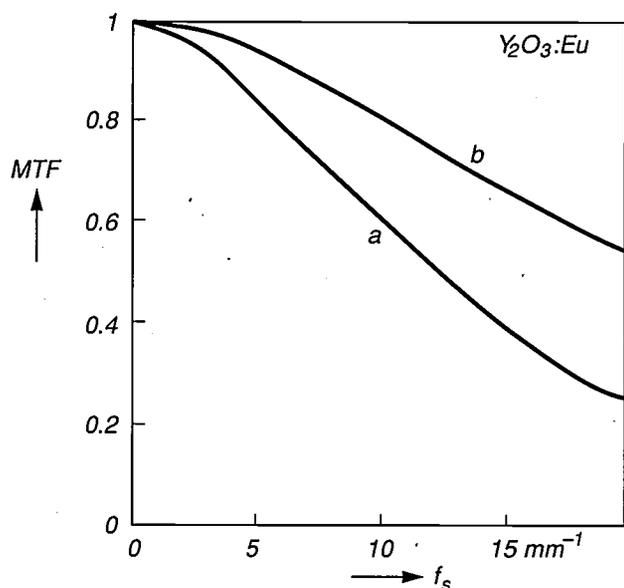


Fig. 11. Resolution as expressed by the modulation transfer function MTF plotted against the spatial frequency f_s , for the two screens of fig. 10. The screen with the high packing density (b) gives a much better resolution than the one with the low packing density (a).

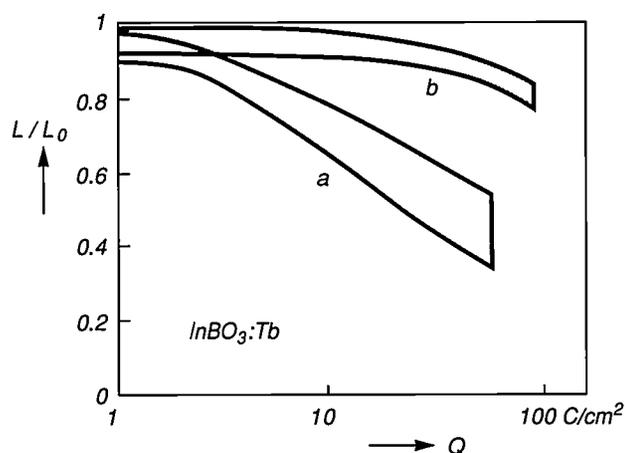


Fig. 12. Effect of the screen processing on the degradation of the light output of a large number of $\text{InBO}_3:\text{Tb}$ powder screens. The ratio of the light outputs after ageing and before ageing (L/L_0) is plotted against the total charge deposited per unit area (Q), a measure of the operating life. Screens prepared by an improved process giving a high packing density (b) degrade more slowly than the screens prepared by a standard process (a).

of two screens containing $\text{Y}_2\text{O}_3:\text{Eu}$ prepared under different conditions. The screen obtained by a standard process has a packing density of between 30% and 40%. An improved process using a buffering electrolyte gives a screen with a packing density of up to 60%. For a given weight this screen gives a much better resolution, because the layer is thinner. This is shown in fig. 11, where the resolution of the two screens is shown as a plot of the modulation transfer function against the spatial frequency.

Preparation processes giving higher packing densities also improve the life. This is shown in fig. 12, where the light output is plotted against the electron charge deposited for a large number of $\text{InBO}_3:\text{Tb}$ screens. A similar improvement has been observed for screens with other phosphors. This has not yet been completely explained, but we can point to two significant factors. One is the low probability of electrons passing through a high-density phosphor layer, so that fewer electrons penetrate the glass substrate to give browning of the glass. The other is related to the difference in binder content. In the standard process, binding is induced by a colloidal silicate, whereas in the improved process the phosphor grains are coated with a dense vitreous silicate layer that stabilizes the phosphor to prevent electron damage. We shall now consider the degradation in more detail.

Degradation

Degradation, i.e. the decrease in light output under prolonged electron bombardment, depends on the phosphor properties, the screen processing and the tube processing. A good way of investigating degradation is by carrying out ageing experiments on actual tubes. We can predict the performance in practical conditions from these experiments, and further analysis on aged tubes will provide a better understanding of the main causes of degradation.

Ageing experiments

Most phosphor screens are subject to 'Coulomb ageing', which means that the decrease in light output at a fixed acceleration voltage depends only on the deposited charge per unit area. An example of such behaviour is shown in fig. 13a for $\text{ZnS}:\text{Ag}$. In this case it is easy to predict the tube performance for different operating conditions and to perform accelerated life tests.

With some phosphors, however, we observe a non-Coulomb ageing: their degradation also depends on the current density during ageing, as shown in fig. 13b for $\text{LaOBr}:\text{Tb}$. This complicates the testing of new phosphors and screen processes. A staircase video pattern dividing the phosphor screen into regions, each aged with a different beam current, has therefore been employed^[10]. This means that the effect of different beam-current densities on a tube can be studied, and the aged screen can also be used for further analysis of the regions that have received different Coulomb doses. Changes in luminescence and optical properties can be measured at regular time intervals during ageing, or they can be measured in different regions after ageing.

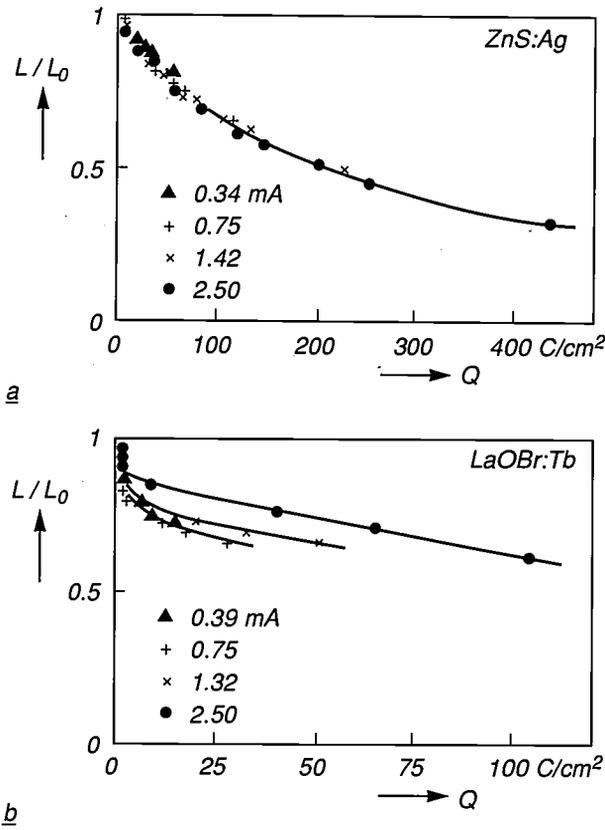


Fig. 13. Relative light output L/L_0 of two types of phosphor screen as a function of the charge deposited per unit area (Q) at different values of the beam current during ageing. The degradation of the light output is independent of the beam current for ZnS:Ag screens (a), but does depend on it for LaOBr:Tb screens (b).

Degradation mechanisms

In discussing possible mechanisms for the decrease in luminous efficacy of the phosphor screen we shall again consider the various stages in the energy-conversion process; see fig. 3 and eq. (1). The parameters sensitive to ageing are the efficiency of the energy transfer from the phosphor host lattice to the activator, the quantum efficiency of the activator luminescence and the photon-escape efficiency of the screen.

The transfer efficiency η_t , can be decreased by generating killer centres in the bulk or at the surface. The presence of these centres can be observed as an increased linearity at high excitation densities. The transfer efficiency can also be decreased by reducing the concentration of effective luminescent centres, which will reduce the light output at high excitation densities [11].

The quantum efficiency η_a of the activator luminescence can be decreased by an increase in the non-radiative loss processes in the activator due to the electron bombardment. The most direct proof of the existence of these processes can be seen in a reduction in the luminescence decay [11].

A decrease in η_t and η_a can also be demonstrated by comparing the light output of aged and non-aged screens as a function of the acceleration voltage. After ageing, a lower output at low voltages (and hence a

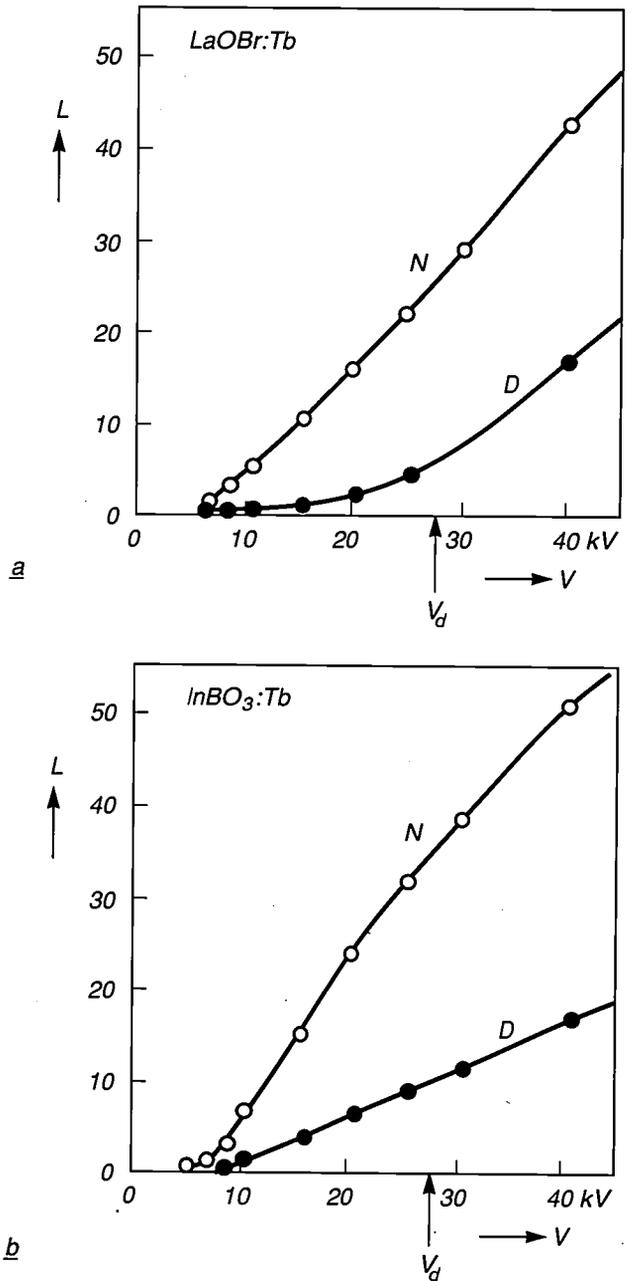


Fig. 14. Light output L (in arbitrary units) measured as a function of the acceleration voltage V for LaOBr:Tb screens (a) and InBO₃:Tb screens (b), non-degraded (N) and degraded (D) at a fixed voltage V_d . In the non-degraded screens the light output increases much more rapidly with voltage. In degraded LaOBr:Tb screens there is a superlinear increase at voltages above V_d .

[10] T. Welker, S. Klauer, J. H. M. Spruit and L. Vriens, Aging of phosphors in projection TV tubes, Extended Abstracts Electrochem. Soc. 87-2 (Fall Meeting, Honolulu, Hawaii, 1987), pp. 1730-1731.

[11] D. B. M. Klaassen, D. M. de Leeuw and T. Welker, Degradation of phosphors under cathode-ray excitation, J. Lumin. 37, 21-28, 1987.

low electron penetration depth) indicates an increase in surface losses. The effect of bulk deterioration on η_t and η_a is revealed by a superlinear increase in the output at higher voltages when the primary electrons excite a non-aged region. Such an increase has been observed in LaOBr:Tb, for example, but not in InBO₃:Tb; see *fig. 14*.

The photon-escape efficiency η_e of the screen depends on the scattering and absorption in the phosphor layer, the reflectance of the aluminium film and the thickness of the phosphor layer. A significant effect that may arise with electron bombardment is an increase in the absorption in the phosphor layer, giving a reduction in η_e . There may also be browning of the glass faceplate; this increases its diffuse reflectance and decreases its transmittance, which also gives a reduction in η_e . The change in η_e can be calculated by the Monte Carlo method from the measured reflectance of the screen and the reflectance and transmittance of the uncoated faceplate.

Analysis of aged screens

A variety of aged screens have been investigated to find the dominant degradation mechanism for various phosphors. We have used the results of the optical measurements to calculate the values of η_e . In *fig. 15* η_e is plotted as a function of the light output, both normalized to the values determined for non-aged screens. The points on the straight line correspond to degradation due entirely to a decrease in η_e . This is the case for ZnS:Ag, LaOCl:Tm, InBO₃:Tb and Y₂O₃:Eu, but not for LaOBr:Tb and Y₂SiO₅:Tb, in which the decrease in η_e is much less than the decrease in light output. In these phosphors a decrease in η_t or η_a , or in both of them, must be the main reason for the deterioration. As would be expected, the variation in their light output as a function of the acceleration voltage gives a superlinear increase at high voltages, as was shown in *fig. 14* for LaOBr:Tb. This gives further evidence of a deterioration in the generation of light (η_t and η_a), since η_e is not very dependent on the position at which the light is generated. We have also observed that these phosphors are more linear after ageing, which indicates that killer centres are formed in the bulk, so that the transfer efficiency is reduced.

Improvement of the degradation behaviour

In phosphor screens whose degradation is dominated by a decrease in the photon escape efficiency there are usually only negligible changes in the luminescence properties. This indicates that increased absorption in the bulk material is unlikely for these 'absorption-deteriorated' phosphors. The absorption is more likely to take place on the surface of the

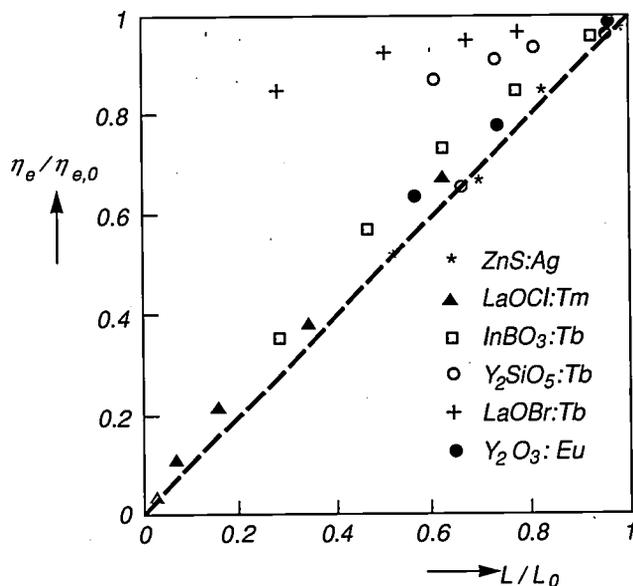


Fig. 15. Calculated values of the relative photon-escape efficiency $\eta_e/\eta_{e,0}$ plotted against the relative light output L/L_0 of various phosphor screens at different stages of degradation. For ZnS:Ag, LaOCl:Tm, InBO₃:Tb and Y₂O₃:Eu the decrease in light output is closely related to the decrease in the escape efficiency. For Y₂SiO₅:Tb and LaOBr:Tb, however, the light output decreases more rapidly than the escape efficiency.

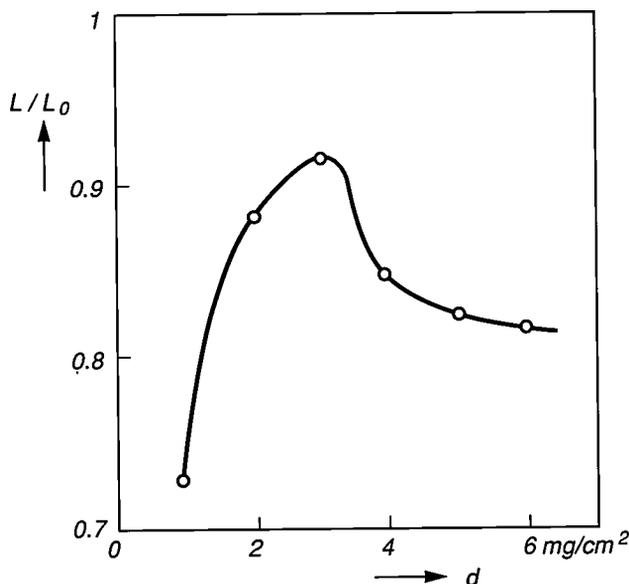


Fig. 16. Typical curve for the relative light output L/L_0 of deteriorated screens as a function of the screen weight per unit area (d). A charge of 100 C/cm² was deposited during ageing. Increasing the screen weight first gives a higher light output, but as the weight increases the light output decreases.

grains, in the binder of the screen or in the browned glass faceplate. The degradation can therefore be reduced by improving the screen processing, as has been demonstrated for screens containing InBO₃:Tb (*fig. 12*).

The degradation of absorption-deteriorated phosphors has been found to depend closely on the screen

weight. A typical example is shown in *fig. 16*, where the relative light output of aged screens is plotted against the screen weight. The increased degradation at low weights can be ascribed to an increased glass browning due to electrons unabsorbed by the phosphor layer. At high screen weights, the degradation increases because of the longer path of the photons before leaving the screen. An increasing degradation due to a larger effective photon pathlength also occurs when these phosphors are combined with interference filters^[1].

A different approach is necessary for 'intrinsic-deteriorated' phosphors such as LaOBr:Tb and Y₂SiO₅:Tb where the reduction in η_t and η_a is dominant (*fig. 15*). An improvement in their degradation behaviour should start with an optimization of the phosphor powder itself. These phosphors are however of interest for use in combination with interference filters, because their degradation is not affected by an increase of the effective photon pathlength.

Outlook

Many aspects of phosphor screens for projection television are now well understood. This is particularly useful for predicting their behaviour in future high-quality large-screen projection systems, in which the improvements will include better resolution, higher luminance and the absence of noise and flicker. It will also be easier to identify any problems that may be associated with such improvements. The use of interference filters between the phosphor layer and the faceplate^[1] can also be taken into account in such predictions; one application of these filters is for adjusting non-ideal chromaticities of new or existing phosphors^[12].

A problem at present is that ZnS:Ag is highly sub-linear. However, there is as yet no fully proven alternative blue-emitting phosphor to compete with ZnS:Ag, even with chromaticity correction from interference filters. The most promising activators from the rare-earth ions are Tm³⁺, which gives a line emission around 460 nm, and Ce³⁺, which gives a blue band emission in suitable host lattices. The best Tm³⁺ phosphor for projection television is probably

La(Ga,Al)O₃:Tm, but its efficacy only approaches that of ZnS:Ag at excitation densities higher than those used in present projection tubes^[13]. A phosphor that can compete with ZnS:Ag at the high densities used in present tubes is (La,Gd)OBr:Ce. However, this phosphor requires a special screen-making process because it is sensitive to water^[14]. The plate-like morphology of the phosphor crystallites also has adverse effects on the photon escape efficiency. Further effort is therefore required if the full potential of these phosphors is to be exploited.

ZnS:Ag can be retained and the effect of sub-linearity can be diminished if the excitation density per pulse is reduced. This can be done by defocusing the electron spot slightly; the defocusing is hardly visible in the picture because the blue resolution of the eye is limited. However, future improvements in picture resolution will require a smaller spot size for blue. A better way of decreasing the excitation density per pulse without losing resolution is to increase the scanning frequency. Quadrupling the scanning frequency would reduce the excitation density per pulse by the same factor, giving much improved linearity for the light output from the tube^[7].

The contribution from the green emission to the luminous output in the white of the picture amounts to about 70%. Improvements in the green-emitting phosphors are therefore always important. A green-emitting phosphor that is also efficient at high densities is LaOBr:Tb, with a luminous efficacy approaching 60 lm/W. However, this phosphor gives much the same problems in screen processing as (La,Gd)OBr:Ce, since it is sensitive to water and has a plate-like grain morphology; further research is required.

It seems unlikely that a replacement for Y₂O₃:Eu as a red emitter will be required in the near future.

The degradation in light output of phosphor screens for projection television is another important area for further research. We have seen that the degradation behaviour of many screens is also related to screen technology, not just to intrinsic phosphor properties. We expect further research to give further significant improvements in degradation behaviour.

[12] D. B. M. Klaassen, D. M. de Leeuw and C. A. H. A. Mutsaers, Projection cathode-ray tubes comprising blue emitting phosphors with interference filters, *J. Electrochem. Soc.* **136**, 858-862, 1989.

[13] K. J. B. M. Nieuwesteeg and C. A. H. A. Mutsaers, Preparation and characterization of thulium-activated La(Al,Ga)O₃ phosphors for blue-emitting cathode ray tubes, *Philips J. Res.* **44**, 157-182, 1989.

[14] D. M. de Leeuw, C. A. H. A. Mutsaers, H. Mulder and D. B. M. Klaassen, Blue emitting phosphors for projection cathode ray tubes, (La,Y)OBr:Ce and (La,Gd)OBr:Ce, *J. Electrochem. Soc.* **135**, 1009-1014, 1988.

Summary. Phosphor screens in tubes for projection television have to meet some critical requirements, since they operate at much higher electron-excitation densities than the screens in conventional direct-view tubes. Their light output should be only slightly sub-linear at high excitation densities and should be stable under prolonged intense electron bombardment. This affects the selection of the phosphors and the screen processing. The preparation of screens with the optimum light output and resolution requires a careful evaluation of their optical properties and degradation behaviour.

Electron guns for projection television

T. G. Spanjer, A. A. van Gorkum and W. M. van Alphen

While the previous article was about phosphors in cathode-ray tubes for projection television, the article below is about the electron guns that excite the phosphors. As in conventional tubes for direct-view television, these guns contain a cathode for electron supply and a number of electron-optical lenses that form the electron beam and focus it on the phosphor screen. To produce the sharp bright images required for projection television, the beam must have a high intensity and a very small diameter at the point where it meets the screen. This sets difficult requirements on the design of the electron guns.

Introduction

The electron gun in a cathode-ray tube for television produces an electron beam that excites the phosphor layer (the 'phosphor screen') on the faceplate of the tube, causing it to emit light. In direct-view colour television this screen contains three different phosphors that emit light in the primary colours red, green and blue, and the viewer looks directly at the picture on the faceplate of the tube. In projection television three small cathode-ray tubes provide separate images in red, green and blue. These images are magnified about 10 times and superimposed on a large projection screen. The pictures should be comparable in brightness and resolution with direct-view television. This requires electron guns that provide a much smaller electron spot at the screen than in conventional cathode-ray tubes. Phosphors are also required with a high light output, which is still reasonably linear with beam current even at high current densities.

In the last few years there has been greatly increased interest in projection television, mainly because improvements in various components have given much better picture quality. A further stimulus to the application of projection television will be the introduction of high-definition television (HDTV) with large screens.

In an earlier article in this journal^[1] it was shown how the brightness, colour rendering and resolution of projection television can be improved by applying interference filters to the faceplates of the three tubes and by using curved faceplates in the tubes. We have already considered the behaviour of phosphors in the previous article^[2]; now we shall consider the electron guns.

The principle of a cathode-ray tube for television and the operation of an electron gun are illustrated in the diagram of *fig. 1*. The triode of the gun emits a divergent electron beam whose intensity at any instant is proportional to the brightness of the pixel to be reproduced. The beam is focused by an electron-optical lens (the 'main lens'), consisting of regions with differing electric (or magnetic) fields, to produce a small 'electron spot' at the phosphor screen. Between the gun and the screen the beam is deflected by two time-dependent magnetic fields to 'write' the image on the phosphor screen. The intensity of the light emitted is determined by the beam current and the luminescence properties of the phosphor layer. The resolution is determined by the size of the electron spot and the light scattering in the phosphor layer.

The tubes used for projection television are small, with a typical usable screen diagonal of 125 mm. To obtain sufficiently bright pictures at the phosphor

Dr Ir T. G. Spanjer, Dr Ir A. A. van Gorkum and Dr W. M. van Alphen are with Philips Research Laboratories, Eindhoven.

screen, the beam current should be high, up to about 6 mA. To give the required resolution, the electron spot should have a diameter no more than a few tenths of a millimetre, much smaller than for conventional direct-view tubes at about the same current (a few millimetres).

The objective of the study described in this article was the design of electron guns that would meet these requirements. One difficulty in reducing the size of the spot is that of aberrations in focusing the beam on to

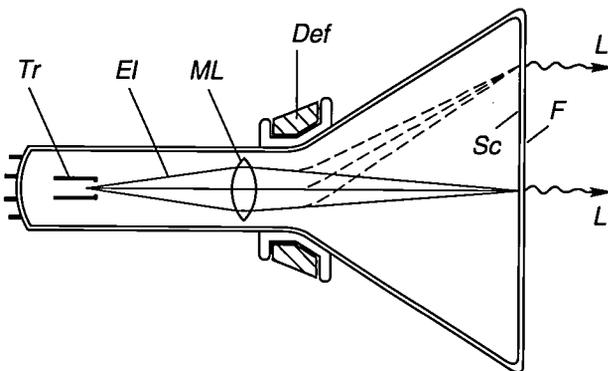


Fig. 1. Principle of a cathode-ray tube for television. The electron gun in this tube consists of a triode *Tr* that gives a divergent electron beam *El*, and a main lens *ML* that focuses the beam on to the phosphor layer of the screen *Sc*. Time-dependent magnetic fields from the deflection coil *Def* deflect the beam in two directions, so that the image is 'written' on the screen. The electrons that strike the phosphor layer excite light *L*, which emerges via the faceplate *F* of the tube.

the screen. These arise because the electrons — or 'electron rays' — at the outside of the beam are deflected too strongly ('spherical aberration'). These errors make a significant contribution to the size of the spot on the screen. The effect of spherical aberration can of course be reduced by reducing the beam diameter in the main lens by using a prefocusing lens, but the improvement is insufficient for high-definition projection television.

After calculating the electron trajectories and the spot size in various kinds of gun, we have been able to derive a new design. This is based on a triode with an impregnated cathode that permits higher beam loading and therefore a more intense beam than a conventional oxide cathode^[3]. Another improvement relates to the main lens: the conventional electrostatic lens, formed by electrodes at different potentials, has been replaced by an electromagnetic lens, formed by a coil around the neck of the tube. This has the advantage that the lens diameter can be larger for given tube dimensions, so that the spherical aberration is reduced^[4].

However, the greatest improvement is obtained by adding two extra electrodes to form a *selective* prefocusing lens. The advantage of using such a lens, also known as an aberration reducing triode, or ART, had already been demonstrated earlier in calculations of the spot size^[5] and in applications in other electron guns^[6]. This lens is positioned so that well before the main lens the electron rays at the outside of the beam are interchanged with rays inside the beam. It considerably reduces the total spherical aberration and gives a more uniform intensity distribution in the beam. This results in a much smaller electron spot.

The new design has been optimized with the aid of electron-optical calculations and measurements. Guns and tubes of this design have been made and the characteristics relevant to projection television have been investigated. We have found that a spot diameter of 0.185 mm can be obtained at a beam current of 4 mA. This means that the brightness and resolution meet the requirements for high-definition projection television.

In this article we shall first look more closely at the resolution of electron guns, and at possible ways of improving them. We shall then give some details of the new electron-optical design and the construction of the gun and tubes. Finally, the most important results from the measurements will be discussed.

The resolution of electron guns

Limitations of the triode

The general configuration and operation of a triode in an electron gun are shown schematically in *fig. 2*. A planar cathode emits electrons under the influence of the accelerating electric field of an electrode at a positive potential with respect to the cathode. An intermediate electrode at a low negative potential provides a decelerating electric field, which confines the electron emission to the central part of the cathode. The electrons are deflected towards the axis of the

[1] L. Vriens, J. A. Clarke and J. H. M. Spruit, Interference filters for projection television tubes, *Philips Tech. Rev.* **44**, 201-210, 1989.

[2] R. Raue, A. T. Vink and T. Welker, Phosphor screens in cathode-ray tubes for projection television, this issue, pp. 335-347.

[3] J. Hasker, J. E. Crombeen and P. A. M. van Dorst, Comment on progress in scandate cathodes, *IEEE Trans.* **ED-36**, 215-219, 1989.

[4] A. A. van Gorkum and T. G. Spanjer, A generalized comparison of spherical aberration of magnetic and electrostatic electron lenses, *Optik* **72**, 134-136, 1986.

[5] A. A. van Gorkum and M. H. L. M. van den Broek, Spot reduction in electron guns using a selective prefocusing lens, *J. Appl. Phys.* **58**, 2902-2908, 1985.

[6] S. Ashizaki, Y. Suzuki, O. Konosu and O. Adachi, 43-inch direct-view color CRT, *Jap. Display* 1986, pp. 226-229; J. Gerritsen and P. G. J. Barten, An electron gun design for flat square 110° color picture tubes, *Proc. SID* **28**, 15-19, 1987.

rotationally symmetric field, thus producing a 'crossover'. The electric field that produces this convergence is called the cathode lens. After the electrons have passed the crossover, they diverge until they reach the main lens, which produces an image of the crossover on the phosphor screen.

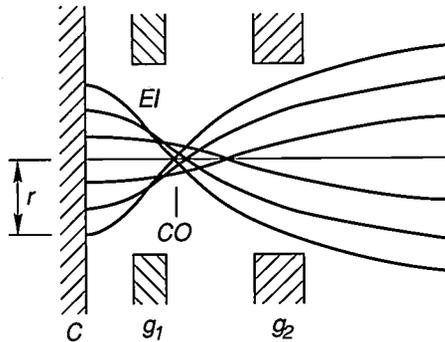


Fig. 2. Schematic geometry and operation of a triode in an electron gun. Electrons EI emerge from the surface of the cathode C . The electrode g_1 has a low negative potential with respect to the cathode, so that the electrons are decelerated and the emission remains limited to an area of radius r . Because of the presence of the electrode g_2 , which has a high positive potential, a 'crossover' CO is produced close to g_1 .

In the ideal case the crossover would be a point. But in practice there are aberrations, which means that the crossover occupies a region around the axis (fig. 2). The most important contribution to the increase in the magnitude of the crossover comes from the spherical aberration of the cathode lens: electrons emerging from the cathode further away from the axis will cross the axis closer to the cathode than electrons starting closer to the axis. Another effect is that the electrons do not all leave the cathode along a normal: the angle of emergence can have any value between 0 and 90° because of the thermal spread in the transverse velocities of the electrons. The contribution from this effect to the size of the crossover depends on the cathode temperature, the type of cathode and the beam current. A third effect is the 'space-charge effect': the electron concentration at the crossover is associated with a strong interaction between the electrons, in which they tend to repel one another. This effect increases with the beam current.

It is not easy to quantify the individual contributions from each of these three effects, since they are all correlated. Nevertheless, it is clear that they degrade the resolution appreciably. The final resolution is degraded even further by focusing errors at the phosphor screen originating from spherical aberration in the main lens.

Spherical aberration of various main lenses

Various kinds of main lens can be used for imaging the crossover on the phosphor screen. We have made calculations to compare the spherical aberration in three different types, one electromagnetic, the other two electrostatic. It was assumed in the calculations that the brightness was constant at the crossover and that the beam diameter in the main lens had the optimum value. We expressed the spherical aberration in terms of three characteristic geometrical parameters. The first parameter is the distance P from the centre of the lens to the crossover. This distance, approximately equal to the length of the electron gun, is one of the factors that determine the length of the tube and is therefore a critical parameter. The second parameter is the distance Q from the centre of the lens to the phosphor screen, which depends on the screen dimensions and the deflection angle. The third parameter is the lens diameter D , which in practice is mainly determined by the diameter of the neck of the tube.

A diagram of the lenses that we have studied is given in fig. 3. In the electromagnetic lens the magnetic field is produced by a coil inside a cylindrical iron yoke. The yoke has an internal diameter of D and a gap of width $0.1 D$. One of the electrostatic lenses is a 'bipotential lens' formed by two cylindrical electrodes, with the same diameter D and a gap width of

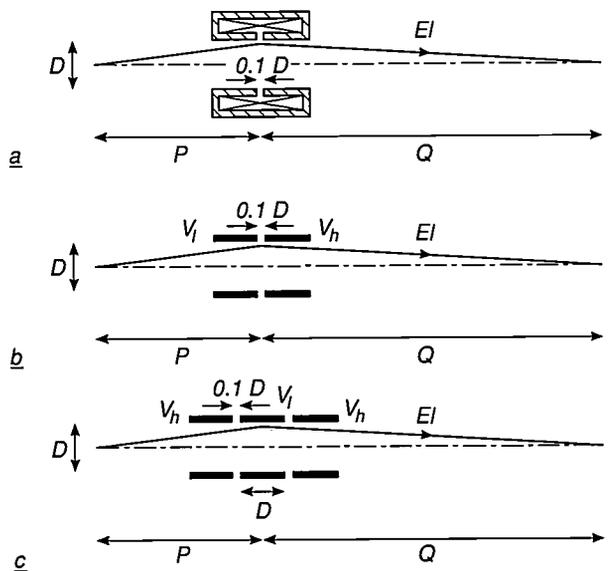


Fig. 3. Geometry of three types of focusing lens for which the spherical aberration has been compared. The important parameters are the lens diameter D and the distances P from the centre of the lens to the crossover and Q from the centre of the lens to the screen. a) Electromagnetic lens formed by a coil in a cylindrical yoke. b) Electrostatic bipotential lens formed by two cylindrical electrodes, one at a low potential V_1 and the other at a high potential V_h . c) Electrostatic unipotential lens formed by three cylindrical electrodes, with the central lens at a potential V_1 and the others at a higher potential V_h .

0.1 D , but at different potentials. The other lens is a 'unipotential lens' formed by three cylindrical electrodes of diameter D and gap widths 0.1 D , with the two outer electrodes at the same high potential and the central electrode, of length D , at a lower potential.

The calculations were performed for guns with P between 50 and 100 mm and Q between 150 and 210 mm, typical values for projection-television tubes. Diameters between 18 and 36 mm were used for the electromagnetic lens in the calculations, and diameters between 12 and 30 mm for the electrostatic lenses. We calculated the quantity $C^{1/4}$ as a measure of the spherical aberration, where C is the aberration coefficient [4]. The value of $C^{1/4}$ is proportional to the smallest spot diameter that can be obtained with the optimum beam diameter in the main lens, at a given brightness of the beam emerging from the triode.

In fig. 4 the quantity $C^{1/4}$ is plotted against P for Q -values of 150 and 210 mm and a lens diameter of 30 mm. This figure shows that the spherical aberration is much the same for each of the three types of lens. It can also be seen that increasing P has little effect: a slight reduction of the spherical aberration. A change in Q has a greater effect: for all three kinds of lens the spherical aberration is much larger for $Q = 210$ mm than for $Q = 150$ mm.

The calculations also show that $C^{1/4}$ is proportional to $D^{-1/2}$. An electromagnetic lens has definite advantages for reducing the spherical aberration by increas-

ing the lens diameter: since the magnetic field is produced outside the neck of the tube, the diameter can be larger for the same tube dimensions than with an electrostatic lens. The results described here were used in optimizing the new electron-optical design.

Improvement by selective prefocusing

To reduce spherical aberration in the focusing action of the main lens a prefocusing lens is often used. This lens is inserted between the crossover and the main lens, so that the beam diameter at the main lens is reduced. The prefocusing lens is usually located some way from the crossover; see fig. 5a. The prefocusing

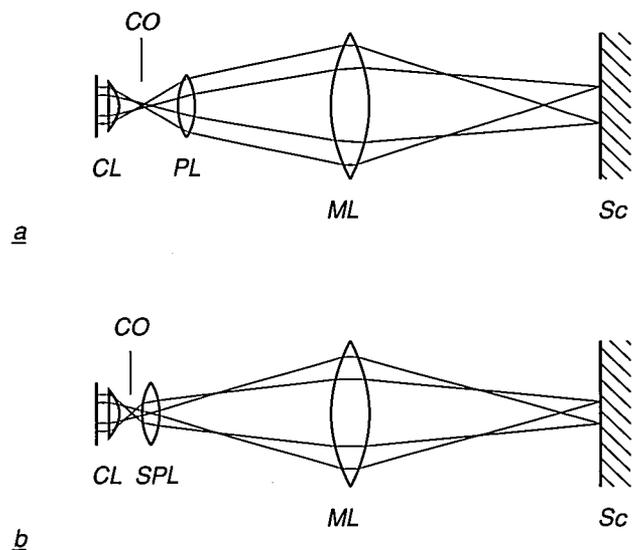


Fig. 5. Effect of the spherical aberration in an electron gun with a conventional prefocusing lens PL between the crossover CO of the cathode lens CL and the main lens ML (a), and with a strong selective prefocusing lens SPL (or aberration-reducing triode, ART) close to the crossover (b). The selective prefocusing forces the outer rays of the beam inwards. This gives a smaller electron spot on the screen Sc when the beam is focused by the main lens, and a more uniform intensity distribution in the beam.

lens performs an integral converging action on all the rays in the beam. The spherical aberrations of the cathode lens and the main lens are usually additive, because the outer rays at the cathode remain on the outside of the electron beam throughout the entire gun. The result is that the spherical aberration is too high for satisfactory application in high-definition projection television, particularly at high beam currents.

A considerable improvement is obtained if the prefocusing lens is made stronger and placed much closer to the crossover; see fig. 5b. The outer rays, which cross the axis some way away from the crossover, are then much more strongly refracted, while the inner rays are hardly refracted at all. This selective prefocusing interchanges the inner and outer rays, so that

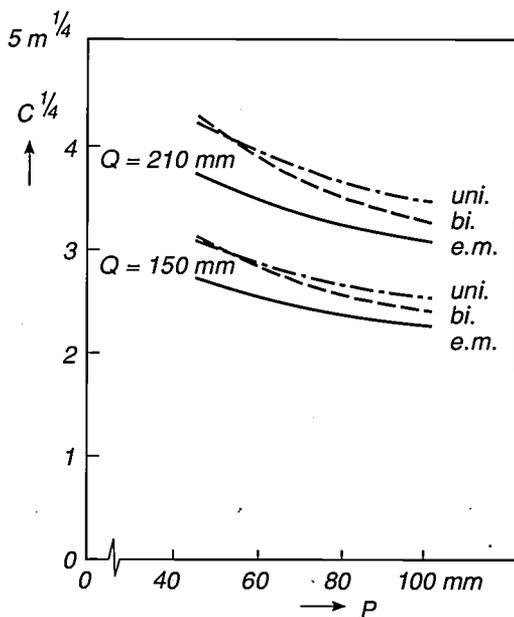


Fig. 4. Calculated value of $C^{1/4}$, a measure of the spherical aberration, as a function of P at two values of Q for the three lenses of fig. 3 with $D = 30$ mm. The electromagnetic lens gives the lowest spherical aberration, but the difference from the other lenses is not particularly great. The effect of Q is larger than that of P .

the spherical aberrations of the cathode lens and the main lens are no longer additive. The result is a very considerable reduction in the total spherical aberration and therefore a much smaller electron spot.

Calculations of the spot diameter show that the effect of the selective prefocusing lens is very dependent on the position and strength of the lens. Some results are shown in *fig. 6*. This gives contours of the normalized spot diameter as a function of the distance from the crossover and the strength of the lens. With the lens very close to the crossover the spot diameter can be reduced to about 55% of the value without prefocusing.

The effect of selective prefocusing can be illustrated by a 'phase-space diagram' for a plane close to the crossover and perpendicular to the tube axis. This is a plot of the angle ϕ between the electron rays and the tube axis against the distance R to the centre, i.e. the intersection of the tube axis with this plane. In the ideal case, where the crossover is a point, all the rays pass through the centre and the phase-space diagram is part of a straight line along the ϕ -axis. In practice, however, most of the rays do not pass through the centre (see *fig. 2*), so that R varies. The magnitude of this variation depends on the size of the crossover and on the plane of the phase-space diagram. For focusing with the main lens we are interested in the plane for which imaging on the phosphor screen gives the smallest spot; this plane is called the object plane of the main lens. This is not the plane for which the diameter of the crossover is a minimum; because of the spherical aberration in the main lens this plane is further away from this lens.

Fig. 7 shows the phase-space diagram of the object plane with no selective prefocusing and with an optimum selective prefocusing lens. With no selective prefocusing the rays with the highest ϕ -values (the outer rays) pass through the centre exactly ($R = 0$), while the rays with half that ϕ -value have the highest R -value. With selective prefocusing the angle of the outer rays is reduced, while the effect on the rays near the centre is small. The largest angles now correspond to the intermediate rays, which means that the inner and outer rays have been interchanged.

The 'acceptance curves' are also shown, to explain the effect on the spot size a little more clearly. These curves indicate the combinations of R and ϕ that correspond, after focusing by the main lens, to the same radius at the screen, proportional to the R -value at $\phi = 0$. From the acceptance curves that enclose the phase-space diagram it is therefore possible to derive a measure for the spot diameter. From the curves given in *fig. 7* it can be seen that selective prefocusing gives a maximum reduction in spot diameter of 46%.

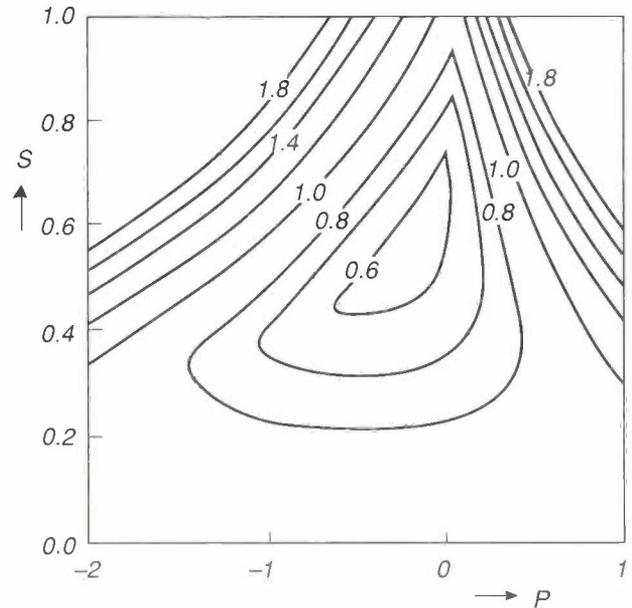


Fig. 6. Calculated contours for the spot diameter, normalized to the value without prefocusing, as a function of the position P of the selective prefocusing lens (with respect to the crossover) and of the strength S of this lens (in arbitrary units). At a position very close to the crossover the spot diameter can be reduced to about 55% of the original value.

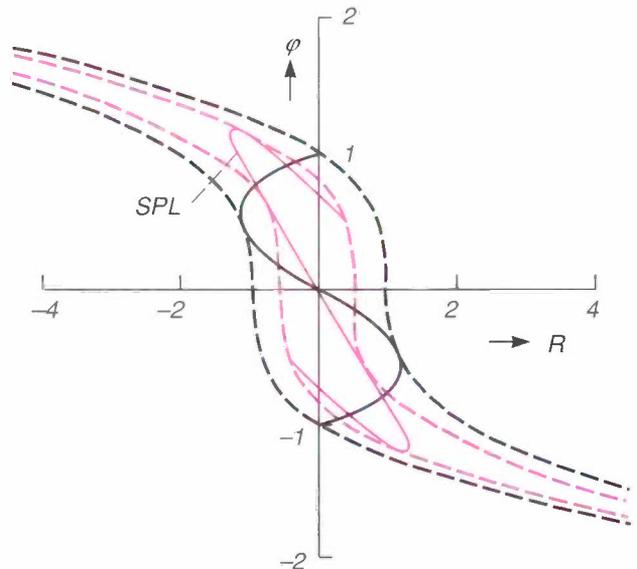


Fig. 7. Calculated phase-space diagrams (continuous lines) for the object plane of the main lens, both with and without an optimum selective prefocusing lens (SPL), and the corresponding acceptance curves (dashed lines). In these diagrams the angle ϕ between the rays and the axis of the tube is plotted against the distance R to the centre of the beam. With the selective prefocusing lens the outer rays (the rays with the highest ϕ) make a smaller angle to the axis and the rays at the centre make a larger angle. The acceptance curves indicate the combinations of R and ϕ that correspond to a particular diameter of the spot at the screen; this diameter is proportional to the distance between the curves at $\phi = 0$. By comparing this distance for the two cases it can be shown that the spot diameter is reduced by 46% with the selective prefocusing lens.

When selective prefocusing is used, the contribution from the space-charge effect to the spot size is also reduced. This is due to a change in the intensity distribution in the beam. In conventional guns this distribution resembles the distribution at the cathode: very inhomogeneous, with a high intensity at the centre of the beam and a very low intensity at the edge. With selective prefocusing the weak outer rays are turned inwards, so that the intensity distribution in the beam becomes more uniform. When the intensity distribution is more uniform the space-charge effect makes a smaller contribution to the spot size.

The new electron-optical design

The beam-forming section

In the new electron-optical design the beam-forming section has been made so versatile that the beam diameter and the position and strength of the selective prefocusing lens can easily be modified by changing the potential of one or more of the electrodes. The design makes full use of the expertise gained in dealing with the crossover, the spherical aberration and the space-charge effect described earlier. The design is also based on the 'cup' model for the cathode lens^[7], which can be used for calculating the characteristics of the crossover as a function of a single geometrical parameter, the ratio of the distance l between the cathode and the end of the first electrode g_1 to the radius r of the emission region. The geometry and potentials required and the nature of the phase-space diagram at the crossover can be derived from the calculations, which take account of the space-charge effect.

The triode in the new design has an impregnated cathode^[3] that gives a maximum current density of 10 A/cm^2 . This means that an emission region with a radius r of 0.2 mm is large enough to give a maximum beam current of 5 mA . The radius of the cylindrical aperture in the first electrode g_1 is also made 0.2 mm , so that the emission region is no greater than this aperture at the maximum beam current. Since the optimum position of the selective prefocusing lens is very close to the crossover, the second electrode g_2 should not have too high a potential and should be close to g_1 . To ensure that sufficient field-strength could nevertheless be obtained at the cathode, we used a very small value, 0.075 mm , for the spacing between g_1 and the cathode and for the thickness of g_1 , so that the geometrical parameter l/r for the cup model is 0.75 . Using this value in the calculations gives a crossover at 0.56 mm from the cathode. This value sets the maximum distance between the cathode and the end of g_2 , since the first electrode of the selec-

tive prefocusing lens must be close to the crossover. The thickness of g_2 and its spacing from g_1 have both been set at 0.15 mm .

Two electrodes are used for selective prefocusing. The first electrode (g_3) is located only 0.25 mm from g_2 and has a low potential. The second electrode (g_4) is located more than 1 mm further away and has a much higher potential, which forms a transition to the high potential (about 30 kV) of yet another electrode, which is connected to the screen, the 'screen electrode' (g_5).

The geometry of the beam-forming section in the new design is shown in the diagram of *fig. 8*. Some

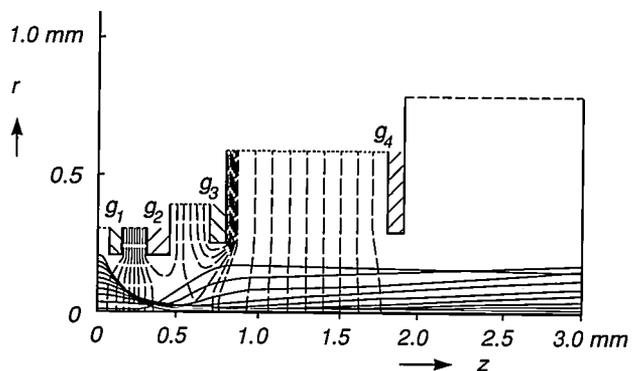


Fig. 8. Geometry of the beam-forming section in the new design of gun, where z is the distance to the cathode and r is the distance to the axis. The electrodes g_1 and g_2 , which determine the formation of the crossover, are very thin and are very closely spaced, and are located a short distance away from the cathode. Close to these is another very thin electrode g_3 . This electrode and the electrode g_4 form the selective prefocusing lens. Dashed lines: calculated equipotentials when the potentials of g_1 , g_2 , g_3 and g_4 are -1 , 500 , 200 and 5000 V respectively. Continuous lines: calculated electron trajectories in this field, with a beam current of about 5 mA . The outer rays of the beam are forced inwards by the selective prefocusing.

equipotential lines are also shown, as well as some calculated trajectories for electrons travelling through the potential field of the electrodes. For clarity the figure only shows electron trajectories that start at right angles to the cathode surface and are unaffected by the space-charge effect. It can be seen from these electron trajectories that selective prefocusing does indeed occur: just beyond the crossover the outer rays are forced towards the inner regions of the beam.

The electron-trajectory data can be more easily analysed by calculating the phase-space diagram for the object plane. The calculated phase-space diagram for six values of the beam current is shown in *fig. 9*. At the highest current the shape of the diagram is very close to the shape required for optimum imaging

[7] A. A. van Gorkum, The cup model for the cathode lens in triode electron guns, *Optik* 71, 93-104, 1985.

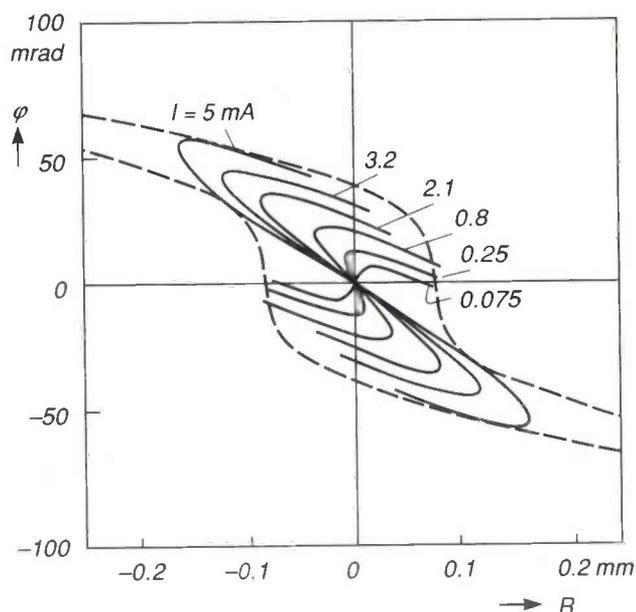


Fig. 9. Calculated phase-space diagrams (continuous lines) for the new design at the object plane of the main lens, for six values of the beam current I , and acceptance curves (dashed lines) for a particular size of spot. The shape of the diagrams depends strongly on I : at lower currents the maximum value of ϕ is much lower than at 5 mA. All the diagrams fit inside the acceptance curves shown except at $I = 0.8$ mA.

(fig. 7). At lower currents this is no longer the case, because of the shift of the crossover with respect to the selective prefocusing lens. At the lowest current some of the rays are in fact focused by this lens: the angle of these rays becomes zero or changes sign. This means that they cross the axis because of the selective prefocusing. Comparison of the phase-space diagrams with the calculated acceptance curves for a particular spot size (see fig. 9) shows whether this affects the imaging. At high currents the phase-space diagram fits inside the acceptance curves, but not when the current is reduced to 0.8 mA. At the two lowest currents the diagram fits inside again. This means that outer rays crossing the axis will not make the spot larger than for the larger currents. A slight increase in spot size as a function of current will occur for currents of about 0.8 mA.

Further calculations have shown that the potential of g_3 has a considerable effect on the shape of the phase-space diagram, especially on the behaviour of the outer rays of the beam. A lower value for this potential gives stronger selective profocusing.

The electrons also pass through the lens formed by g_4 and the screen electrode g_5 . In the new design electrons travelling from g_4 to g_5 are exposed to the effect of a weak negative lens. The strength of this lens, and hence the diameter of the beam, can be varied by altering the potential of g_4 .

The main lens

For the main lens in the new design we decided to use an electromagnetic lens, of the type shown in fig. 3a. Our choice was based on the results mentioned earlier of the calculations of the spherical aberration with various kinds of lens. The calculations showed that for a given size of tube an electromagnetic lens can have a smaller spherical aberration than electrostatic lenses, since its diameter can be larger.

A description of the geometry and operation of the main lens can be found in fig. 1 and fig. 3a. The cylindrical yoke for magnetic focusing has an internal diameter of 50 mm and a gap with a width of 10 mm. The centre of the lens is located 95 mm from the cathode and 175 mm from the inside of the phosphor screen.

Construction

An important feature of the design is that the electrodes g_1 , g_2 and g_3 are very thin and very closely spaced (see fig. 8). In making the experimental guns special construction techniques were necessary to produce this kind of configuration. We used a method in which the electrodes were soldered to partly metallized sapphire rods. The very small spacings required can be produced and maintained with this method.

The rest of the gun was manufactured by conventional methods, with the electrodes held in position by glass insulator rods. A photograph of the gun is shown in fig. 10. The electrodes g_1 , g_2 and g_3 cannot be separately distinguished in the photograph. The electrodes g_4 and g_5 , the glass rods that support them and the first part of the connection to the screen are clearly visible, however.

The cathode-ray tubes that have been made have a screen diagonal of 125 mm, a deflection angle of 55° and a neck diameter of 36 mm. In most of the experiments the screen was coated with a fine-grain green-

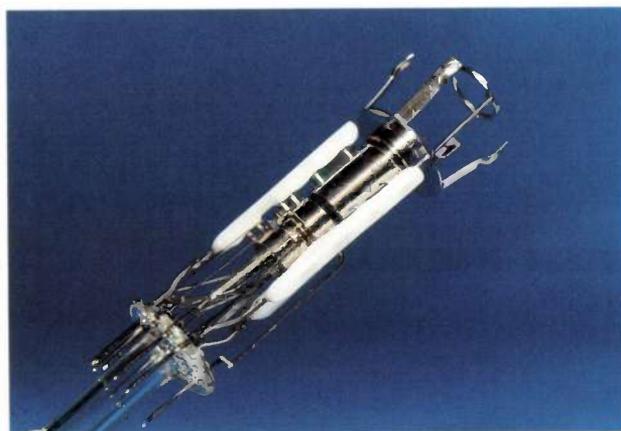


Fig. 10. The new design for the electron gun.



Fig. 11. Cathode-ray tube with the electron gun of fig. 10. The length of the tube is about 300 mm; the neck diameter is 36 mm and the screen diagonal is 125 mm.

emitting phosphor (LaOBr:Tb) applied at high packing density to reduce light scattering in the phosphor layer. A photograph of the tube, also showing the electromagnetic main lens, is shown in fig. 11.

Measurements

We have made a large number of measurements on both the guns and the completed tubes to determine the characteristics of interest for projection television. The measurements on the guns were mainly concerned with the phase-space diagrams for the beam. The spot size for the tubes was determined, as well as the picture resolution obtained. The measurements were made with various combinations of electrode poten-

tials between these electrodes. The diagrams were measured for various values of the beam current and the potential of g_3 . Some results obtained at the optimum value of the potential of g_3 and three values of the beam current are shown in fig. 12. The shape of the diagrams is very dependent on the beam current, as would be expected from the results of the calculations (fig. 9). At 3 mA the diagram has the shape required for optimum selective prefocusing. At 1 mA the outer rays are pulled inwards towards the centre of the beam; some of these rays are almost parallel to the axis of the tube ($\phi = 0$). At the smallest current (0.2 mA) the diagram looks quite different; an intermediate crossover is formed by the selective prefocusing.

The diagrams for the object plane of the main lens can be derived from the measured phase-space diagrams. The object-plane diagrams are in good agreement with the calculated diagrams shown in fig. 9. These results confirm that the introduction of a variable lens close to the crossover does indeed provide the desired selective prefocusing. It can also be seen that this lens is strong enough to modify the phase-space diagrams in the appropriate way.

Resolution

The resolution of the picture is determined by the diameter of the electron spot at the screen. We have measured this at various values of the electrode potentials and beam current, and also as a function of the

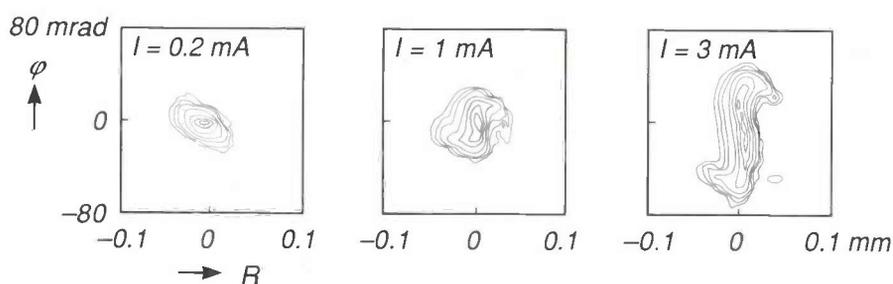


Fig. 12. Measured phase-space diagrams for optimum selective prefocusing, for three values of the beam current I . The contours shown correspond to 2, 5, 10, 25, 50, 75 and 90% of the maximum intensity at the centre. The measured change in shape with beam current corresponds to the results of the calculations.

tials and currents. We shall now discuss the most important results briefly.

Phase-space diagrams

Detailed phase-space diagrams for the beam as it emerges from the beam-forming section were measured in a special system^[8]. Because of experimental limitations, the same value (7 kV) was used for the potentials of g_4 and g_5 , so that there was no lens action

focusing current, i.e. the current in the coil of the electromagnetic main lens. We have determined the spot diameter as the distance between the points where the intensity is 5% of the maximum value. From the measurements at various values of the potentials of g_3 and g_4 it was possible to derive the combination of potentials that gives the smallest spot diameter.

[8] M. H. L. M. van den Broek, Experimental emittance diagrams of triode electron guns, *J. Phys. D* 19, 1401-1419, 1986.

Fig. 13 shows the variation as a function of the focusing current for this optimum combination, at various values of the beam current. The focusing current at which the smallest spot is obtained is only slightly dependent on the beam current. This means that there

are no undesirable effects if the beam current is fixed, as in normal practice.

Under optimum conditions with a fixed focusing current the measured spot diameter varies very little with beam current; see fig. 14. As would be expected from the phase-space diagrams, the measured spot diameter at low currents is rather smaller and there is a slight enlargement at about 0.8 mA. At currents up to 4 mA the measured spot diameter remains smaller than 0.210 mm. In practice the spot diameter is even less, since the light scattering in the phosphor layer also contributes to the spot size in these measurements. A rough estimate of this contribution indicates that the true spot diameter is no larger than 0.185 mm, and a very sharp image of the crossover is therefore produced.

This has been confirmed by determining the 'modulation transfer function' (or MTF). The MTF represents the fraction of the amplitude transmitted for a sinusoidal signal at a particular spatial frequency. (The spatial frequency is the number of periods per unit length.) Fig. 15 shows the MTF-value as a function of the spatial frequency for an optimum tube, for four values of the beam current and a fixed focusing current. In assessing these results we have to remember that for a standard PAL television picture on a screen with a 125-mm diagonal a spatial frequency of 2.5 periods per mm corresponds to a signal at 5 MHz. It can be seen that even at a current of 4 mA twice as much horizontal information (5 periods per mm) can be displayed with an MTF of about 40%. This value of the MTF is more than adequate for a good picture.

To sum up, electron guns based on the new design will give cathode-ray tubes with a very high resolution. Tubes that have been made give pictures whose brightness and resolution are sufficient for projection television and high-definition television (HDTV) with large screens.

Important contributions to the work described here were made by M. R. T. Smits and T. L. van Soest. The guns and tubes were made by F. A. M. Habraken and other colleagues in the Tube, Glass and Cathode department. The phosphor coatings in the tubes were applied by colleagues at Philips GmbH Forschungslaboratorium Aachen.

Summary. Electron guns of a new design have been made for application in cathode-ray tubes for projection television. These guns contain an impregnated cathode, a selective prefocusing lens and an electromagnetic main lens. The electron spot on the phosphor screen can be very small, even at high beam currents (0.185 mm at 4 mA), mainly because of selective prefocusing. This means that the brightness and resolution will be sufficient for high-definition projection television.

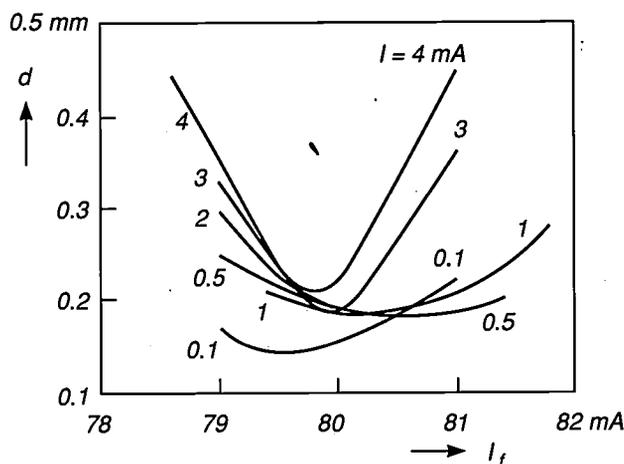


Fig. 13. Measured spot diameter d (distance between the points at 5% of the maximum intensity) as a function of the focusing current I_f , for various values of the beam current I . The beam current does not have much effect on the minimum of the curves.

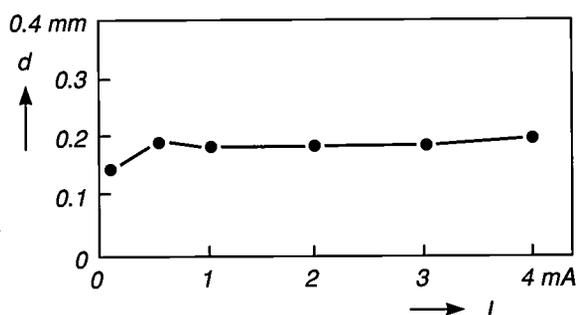


Fig. 14. The spot diameter d measured at a fixed focusing current, as a function of the beam current I . The slightly smaller diameter at the lowest current and the small increase at 0.8 mA are in good agreement with the calculated phase-space diagrams of fig. 9. Even at the highest current the spot diameter is smaller than 0.210 mm.

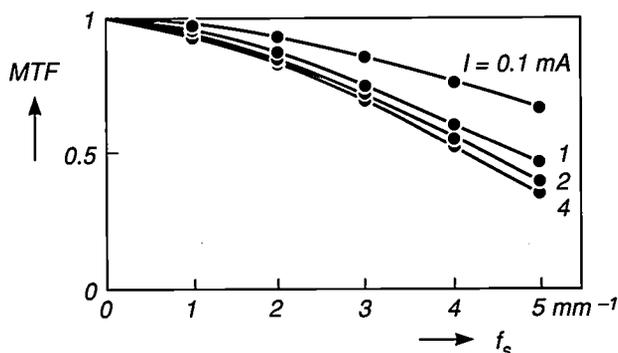


Fig. 15. Modulation transfer function MTF plotted against the spatial frequency f_s (the spatial frequency is the number of periods per mm) at four values of the beam current I . The MTF, a measure of picture resolution, decreases more rapidly with frequency as the beam current increases. At a current of 4 mA and a spatial frequency of 5 mm⁻¹ the MTF is still about 40%.

A diagnostic X-ray tube with spiral-groove bearings

E. A. Muijderman, C. D. Roelandse, A. Vetter and P. Schreiber

The technological development of a product usually advances in a series of small steps. But not always — one example is the change from fixed to rotating anodes in diagnostic X-ray tubes in about 1930. This was a development in which Philips played a leading part, largely because of the work of Albert Bouwers^[1]. Since then there has been a steady stream of small improvements in diagnostic X-ray tubes, mostly because problems with limited bearing life and inadequate cooling of the anode have led designers to make modifications. It now looks as though there has been another great leap forward in solving these problems: spiral groove anode bearings lubricated by liquid metal.

Introduction

A diagnostic X-ray tube has to deliver a large amount of radiation from the smallest possible area of the anode, the 'point focus', in a time of several milliseconds to several seconds^[1]. The quantity of radiation must be large enough for a sufficient number of electrons to be released at the input screen of an X-ray intensifier tube^[2], or to produce sufficient blackening in an X-ray-sensitive film after development. The efficiency of the conversion of electrical energy into X-radiation is less than 1%. More than 99% of the energy supplied to the tube is therefore released in the anode material as heat. In a stationary anode the material would melt locally because the heat is concentrated at the point focus. The problem can be solved by making the anode rotate, since the heat is then spread out over an annular area.

In Philips diagnostic X-ray tubes the surface of the rotating anodes is now made of tungsten alloyed with rhenium. Since the melting point of this alloy is above 3000 °C, high thermal loads at the anode are possible. Alloying with rhenium prevents hairline cracks.

So that the anode can rotate it must be provided with bearings and driven by an electric motor, which has to be an induction motor. The anode, the bearings and the rotor of the induction motor are located inside the evacuated tube. The bearings therefore have to meet a number of rather special requirements. The lubricant for the bearings must not contaminate the vacuum of the tube, the bearings must be able to stand up to high temperatures and they must be electrically and if possible thermally conductive. Solid-metal lubricated ball bearings were therefore always used. The lubricant now used in the high-load Philips diagnostic tubes is silver or lead, applied to the bearing components by special techniques.

Although diagnostic tubes have been produced in this way in large numbers for years, and the technique is still in use, considerable effort is being deployed in the search for alternative methods of providing bear-

Dr Ir E. A. Muijderman was formerly with Philips Research Laboratories, Eindhoven; C. D. Roelandse is with Philips Research Laboratories, Eindhoven; Dipl.-Phys. A. Vetter and Dr P. Schreiber are with Philips Medizin Systeme GmbH, Hamburg.

^[1] A. Bouwers, Eine Metallröntgenröhre mit drehbarer Anode, Verh. D. Röntgen-Ges. Vol. 20, suppl. to Fortschr. Röntgenstr. 40, 102-106, 1929;

J. H. van der Tuuk, X-ray tubes with rotating anode (Rotalix tubes), Philips Tech. Rev. 8, 31-41, 1946.

^[1] W. Hartl, D. Peter and K. Reiber, A metal/ceramic diagnostic X-ray tube, Philips Tech. Rev. 41, 126-134, 1983/84.

^[2] B. van der Eijk and W. Köhl, An X-ray image intensifier with large input format, Philips Tech. Rev. 41, 137-148, 1983/84.

ings for the anode. The main reason for this is the limited life of the ball bearings, which operate in vacuum under extreme conditions, with forces even higher than those in the bearings in the stabilizer flywheels used in space vehicles [3], which also have to operate in vacuum. Another problem is that the usual methods for calculating the life of ball bearings can only be used if these bearings are grease or oil lubricated. Both are not possible in vacuum.

Because of the limited life of the ball bearings the anode is only rotated during the actual exposure. The anode is therefore run up to the maximum speed very quickly: within one to three seconds. The maximum speed will be about 3000 or 9000 revolutions per minute, depending on the tube type, at a mains frequency of 50 Hz. Once the maximum speed has been reached the actual exposure is made; a high electrical current now flows briefly and heat is developed. The anode is then decelerated very rapidly. The electric motor must be capable of developing a very large driving or decelerating torque, because the anode has a high moment of inertia and has to be accelerated and decelerated very quickly. The 'air gap' between rotor and stator is partly outside the actual tube and is wider than 10 mm because of the high voltage between rotor and stator. This means that the electric motor also has to meet some rather special requirements [1].

Accelerating and decelerating the anode for every exposure means that radiologist and patient are forced to accept delays, which are inconvenient for both and may be uncomfortable for the patient. The time required for accelerating the anode is also a problem if the exposure has to be made at a particular phase of a periodic movement, such as the peristaltic movement of the alimentary tract.

Another disadvantage of ball bearings is that they are noisy, since they contain no grease or oil to attenuate the sound. An even greater disadvantage is that ball bearings present a high thermal resistance to the heat to be conducted away from the anode to the outside world. The heat flow that leaves the tube by conduction through ball bearings is therefore virtually negligible, and amounts to only about 3 W. This means that the rest of the heat flow must leave the anode by radiation, which results in a large increase in the mean anode temperature. The last disadvantage of ball bearings that we shall mention is that they present a varying resistance to the anode current.

Another possibility is to use magnetic bearings [4] for the anode. The advantage here is that no lubricant is necessary, since the shaft is 'floating'. However, the equilibrium of the floating shaft is unstable, so that a comprehensive — and therefore expensive — elec-

tronic control circuit is necessary. Since such a circuit can fail (e.g. if the electricity supply fails), which may cause damage, special ball bearings have to be provided that can take over from the magnetic bearings. If these ball bearings are suddenly brought into service they will be subjected to a large acceleration, so that their life is very limited. In practice, this means that the control electronics for the magnetic bearings should never be allowed to fail. Another difficulty with magnetic bearings is that there is no possibility of removing heat by conduction. Also, 'slip rings with



Fig. 1. The new Philips MRC200 diagnostic X-ray tube with an anode mounted in spiral-groove bearings lubricated by liquid metal. The anode diameter is 200 mm. The rated input is 85 kW^[7].

carbon brushes' have to be used for the anode-current supply. (It is also possible to use the principle of the thermionic diode for the passage of current from the stationary tube wall to the rotating anode shaft.) Because of these problems — and also because of the complicated construction of the magnetic bearing — no diagnostic X-ray tubes with magnetic bearings have yet appeared on the market.

At Philips Research Laboratories in Eindhoven another solution has been found to the problem of providing bearings for the rotating anode of a diagnostic X-ray tube: spiral-groove bearings with liquid metal as the lubricant. The application of these bearings solves all of the problems mentioned above:

- life,
- noise,
- heat transfer and
- current supply.

The heat conduction of a metal-lubricated spiral-groove bearing is about a thousand times greater than

that of a ball bearing. Another extremely important advantage is that an anode with spiral-groove bearings can rotate continuously, since its life is virtually unlimited. Periodic acceleration and deceleration are then no longer necessary. This means that radiologist and patient will no longer be subject to delays and that the motor can be much simpler, since it only has to supply enough power to keep the anode rotating continuously. Moreover, a spiral-groove bearing can support a much greater load than a ball bearing, so that an anode of much greater diameter can be used. Since the anode only has to be run up to speed once each day, in the morning, the higher moment of inertia is not a problem.

A spiral-groove bearing is a self-acting bearing, i.e. a bearing that requires no external pressure source, in which there are grooves in one of the two bearing surfaces^[5]. If the relative movement of the bearing surfaces is in the correct direction, the grooves ensure that the lubricant is retained in the bearing. Since the space between the two cylindrical surfaces is wedge-shaped and because of the propulsive action of the grooves an excess pressure is produced in the lubricant. If the excess pressure is high enough, the lubricant will keep the bearing surfaces apart. The principle of the spiral-groove bearing has been known for twenty or thirty years and various forms of these bearings have been studied in depth^[6] at Philips Research Laboratories.

For the application of a spiral-groove bearing in a diagnostic X-ray tube it was necessary to find an alloy that is liquid at or near room temperature and can be used as a lubricant. The alloy must also have a low vapour pressure at high temperatures, so that the vacuum in the tube is not contaminated by vaporized metal. Mercury is therefore unsuitable as a lubricant in vacuum. After extensive experiments by J. Gerkema and J. B. Pelzer at Philips Research Laboratories a gallium alloy was eventually found that can act as a lubricant.

The study at Philips Research Laboratories laid the foundations for a diagnostic X-ray tube that has recently been introduced by Philips Medizin Systeme GmbH of Hamburg. The new tube is shown in *fig. 1*. The type designation MRC 200 is related to the external diameter of the anode, 200 mm. In the present version the anode rotates at almost 3000 rev/min. The rated input power of the tube is 85 kW^[7]. The metal-lubricated spiral-groove bearings have been put through extensive life tests at Philips Research Laboratories. A preproduction batch of tubes has been produced at Hamburg, and these have been thoroughly tested.

The new tube was specially designed for making cine exposures. These are of interest in the investiga-



Fig. 2. Earlier generations of Philips diagnostic X-ray tubes. *a*) The SRO90 tube, *b*) the SRM 100 tube and *c*) the SRC 120 tube. The numbers refer to the anode diameter in mm.

- [3] J. P. Reinhoudt, A flywheel for stabilizing space vehicles, Philips Tech. Rev. 30, 2-6, 1969;
J. Crucq, The reaction wheels of the Netherlands satellite ANS, Philips Tech. Rev. 34, 106-111, 1974.
- [4] E. M. H. Kamerbeek, Magnetic bearings, Philips Tech. Rev. 41, 348-361, 1983/84.
- [5] Strictly speaking the journal bearings are *helical-groove* bearings, but the term 'spiral-groove bearing' is in fact widely used for both spiral-groove thrust bearings and helical-groove journal bearings.
- [6] E. A. Muijderman, New forms of bearing: the gas and the spiral groove bearing, Philips Tech. Rev. 25, 253-274, 1963/64;
E. A. Muijderman, G. Remmers and L. P. M. Tielemans, Grease-lubricated spiral-groove bearings, Philips Tech. Rev. 39, 184-198, 1980.
- [7] The rated input power is defined as the power that can be supplied for a maximum of 0.1 second to a tube that was in thermal equilibrium at a continuous power input of 250 W.

tion of bloodvessels where the blood flow is very rapid, as in the coronary arteries (coronary angiography). At a relatively low radiation dose for each exposure, 6 to 10 series of exposures, each consisting of 300 to 600 exposures, are made for each patient at a rate of 25 to 100 exposures per second. Until recently it was necessary to wait for a fairly long time between the series of exposures until the heat stored in the anode had been dissipated by radiation. The great advantage of an anode with a large diameter combined with metal-lubricated spiral-groove bearings is not only that a large amount of heat can be stored in the anode, but also that this heat can be removed by conduction. The radiologist is therefore no longer forced to wait for the anode to cool down during an investigation. This means that the patient does not have to be subjected to the investigation procedures for such a long time. This can be a great advantage, especially where catheters are used, since more power is then required from the tube. Nor is it necessary to use so much contrast agent. An important economic advantage, of course, is that more patients can be dealt with in a given time.

Besides the version for cine exposures, other versions of the new tube will also be developed for other applications where the excellent cooling and the continuously rotating anode will be advantages. Another advantageous feature of the new tube is the spatial stability of the focus. Because the temperature remains low, the position of the focus never changes by more than 50 μm . The stability of the focus is of interest for the digital processing of X-ray images. With the 'subtraction' method, for example, bloodvessels can be made visible by 'taking away' an image made without a contrast agent from one made with a contrast agent.

Fig. 2 shows the earlier types of tube with rotating anodes: the SRO 90, with glass envelope, the SRM 100, with an envelope of glass and metal, and the SRC 120, made entirely from metal and ceramics^[1]. The numbers in these designations indicate the diameter of the anode in mm; the speed of rotation is almost 9000 rev/min for all the tubes. The introduction of each new tube meant a step in the improvement of the performance. It can be seen from fig. 3 that the introduction of the newest tube corresponded to a much larger advance than the previous ones. The bar chart shows the relative value of various quantities, with the SRM 100 glass/metal tube as the basis for comparison. These quantities are the rated input power^[7], the maximum load during a single cine pulse, the product of heat capacity and mean anode temperature, the quantity of heat that can be stored in the anode during a comparable series of cine exposures (lasting 6 to 10 seconds), the total heat loss due to radiation and conduction,

and the total energy that can be supplied to the tube during a cine investigation lasting 20 minutes. The bar chart shows that this last figure is more than three times higher than for earlier tube types.

We shall now look at the construction of the tube. Next we shall consider the anode bearings, and finally we shall briefly touch on certain aspects of the dynamic stability of the rotating anode.

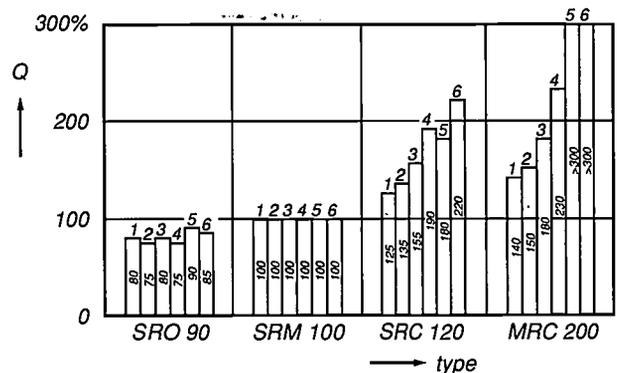


Fig. 3. Bar chart giving a comparison of the performance of successive generations of diagnostic X-ray tubes, see also figs 1 and 2. The relative values Q of a number of quantities are plotted for four different tubes, with Q for the SRM 100 tube given a value of 100 in each case. The values quoted for each type of tube are: 1 the rated input power^[7], 2 the maximum load during a single cine pulse, 3 the product of heat capacity and mean anode temperature, 4 the quantity of heat that can be stored in the anode during a comparable series of cine exposures (lasting 6 to 10 seconds), 5 the total thermal energy that can be dissipated by radiation and conduction, and 6 the total energy that can be supplied to the tube during a cine investigation lasting 20 minutes.

The construction of the tube

Fig. 4 shows a cross-section of the new MRC 200 diagnostic X-ray tube in its shield. The space between tube and shield is filled with oil, which provides high-voltage insulation. The oil also takes part in the cooling process by taking up the heat leaving the anode (A) by radiation and conduction through the spiral-groove bearing (SGB); oil pumped through the tube O_1 is directed against the bottom of the opening in the bearing.

The electrons emitted by the filament of the cathode (K) strike the anode at the point focus (PF). The X-radiation produced here is confined to a conical beam (X) by diaphragms (not shown). The X-radiation leaves the tube through an extremely thin beryllium window (W), a thin layer of oil and a thin aluminium window (Al) in the shield.

The difference between the temperature at the point focus and the mean anode temperature is approx-

imately inversely proportional to \sqrt{nD} , where n is the speed of rotation and D is the outer diameter of the anode^[1]. In designing an X-ray tube with a rotating anode it is necessary to find the right compromise between speed of rotation and anode diameter. In this tube, mainly intended for cine exposures, it was decided to make the speed fairly low and the anode diameter large. In earlier tubes a large anode diameter was undesirable, since running up to speed and decelerating would have taken too long because of the high moment of inertia. In the new tube this is no problem, since it only happens once in each working day. The anode of the MRC 200 therefore has both a high moment of inertia and a high heat capacity. This means that a large amount of heat can be stored tem-

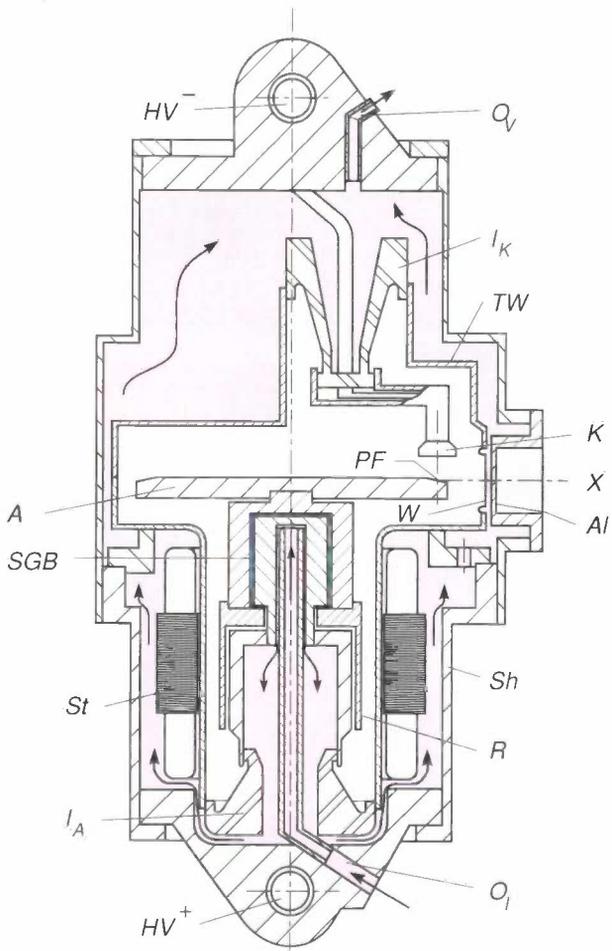


Fig. 4. Schematic cross-section of the MRC 200 diagnostic X-ray tube. White: vacuum. Red: oil; the direction of flow is indicated by arrows. HV^+ positive high-voltage connector. I_A ceramic insulator for the anode. St stator of the induction motor that rotates the anode. SGB spiral-groove bearing to take up forces in both radial and axial directions. (The liquid metal is shown as a blue line.) A anode. HV^- negative high-voltage connector. O_v oil outlet. I_k ceramic insulator for the cathode. TW metal envelope. W beryllium window. K cathode. X X-ray beam (the diaphragms that confine the beam are not shown). Al aluminium window. Sh shield. R rotor of the induction motor. O_i oil supply. PF point focus.

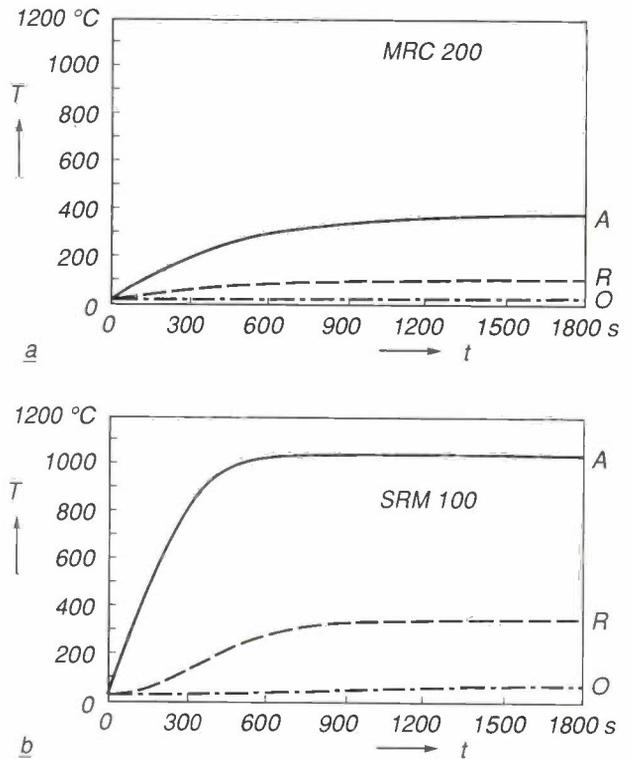


Fig. 5. The temperature T as a function of time t for continuous operation at an input power of 1 kW, starting at time $t=0$, for a) the new MRC 200 tube and b) the earlier type, the SRM 100. A the temperature variation at the anode circumference, R at the rotor circumference and O of the oil in the shield.

porarily in the neighbourhood of the circular path of the point focus. This heat can then be dissipated quickly by conduction through the bearing.

It can be seen from *fig. 5* that the heat capacity of the anode and the heat dissipation by conduction are significantly better than for the earlier SRM 100 tube. The figure shows the results of a computer simulation of the temperature variation in the anode after starting at a continuous tube loading of 1 kW with a relatively large focus. The three curves relate to the temperatures of the anode rim, the rotor and the oil in the shield. All the temperatures for the MRC 200 are considerably lower than for the SRM 100. Since the temperature increase at the anode rim is so much smaller in comparable circumstances, the energy load at the focus can be much higher.

The new tube, like its predecessor, is made entirely of metal and ceramics, and contains no glass at all. Ceramic components can be more accurately dimensioned. The metal envelope is at earth potential. This avoids difficulties from the charge effects that occur in tubes that contain glass in the envelope. The highest potential difference across either of the ceramic insulators (I_k at the cathode and I_A at the anode, see

fig. 4) is about 75 kV, half the maximum voltage across the tube. The high-voltage connectors, HV^+ and HV^- , are located at the two ends of the shield. The connectors are connected to the spiral-groove bearing and the filament of the cathode respectively, and are insulated from the shield by plastic insulators.

As noted earlier, the rotor (R) of the induction motor and the spiral-groove bearing are located in the vacuum. The stator (St) is located outside the tube in the oil between the tube and the shield. Since the anode rotates continuously, the maximum motor power does not have to be much higher than is necessary to overcome the friction in the spiral-groove bearings. The friction loss is only about 40 W.

The anode bearings

The anode spiral-groove bearing, SGB in fig. 4, is constructed as a bearing with a stationary shaft and a rotating bearing bush. Since it has to be possible to use the X-ray tube in all possible orientations, the spiral-groove bearing system consists of two thrust bearings and two journal bearings^[6]; see fig. 6. The pumping action of the different components cancels out, so that no lubricant is forced out of the bearing. This is achieved by using herringbone groove patterns, which give the greatest build-up of pressure at the centre of each pattern. Since there are groove patterns for thrust-bearing action on both upper and lower sides, axial forces can be taken up in two directions. The two separate journal bearings with their own helical groove patterns are necessary for taking up torques as well as forces.

Fig. 6 also shows a diagram of the 'pressure hills' of the different parts of the bearing: the pressure in the lubricant as a function of position. This figure holds for the bearing system in a centric position. When the bearing positions itself eccentrically because of the radial load, the height of the pressure hills is also a function of the angle in the circumferential direction. The peak of the pressure hill is found just in front of the position where the bearing gap is smallest. Integrating each pressure hill gives a part of the bearing force. In principle it makes no difference to the operation of a spiral-groove thrust bearing whether the grooves are in the stationary part of the bearing or in the moving part. In fact we put the grooves in the moving part of the bearing to facilitate production.

Dynamic stability of the rotating anode

An investigation has also been made at Philips Research Laboratories into the possibility of instability of the rotating anode because of a whirling motion

of the axis of symmetry. This 'half-omega whirl' has an angular frequency about half that of the rotational speed. Although this instability can be investigated by using a computer program^[8], results obtained in this way do little to improve the physical understanding of the problem. This means that the search for the cor-

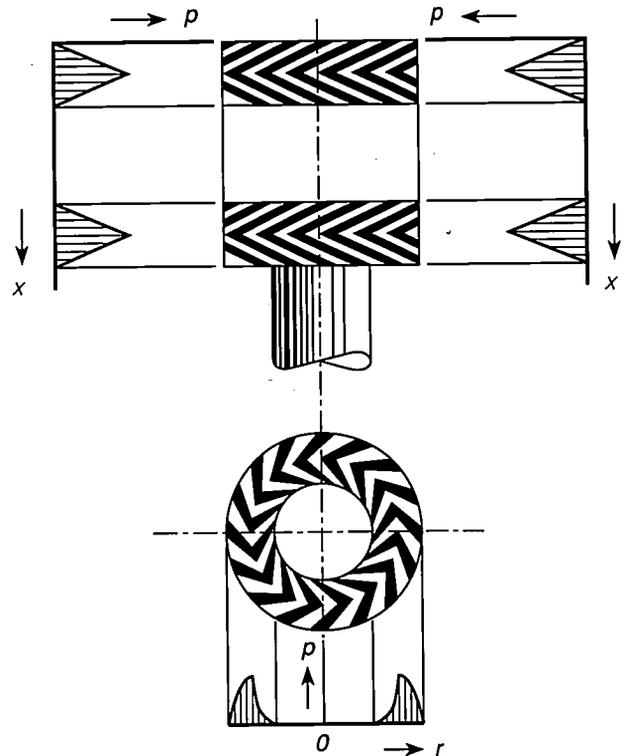


Fig. 6. The spiral-groove bearing system. *Above*: the helical groove patterns of the journal bearings. *Below*: the spiral groove pattern in one of the thrust bearings. The same groove pattern is used in the other thrust bearing (mirror symmetry). The hatched areas are 'pressure hills', representing the pressure p in the lubricant, as a function of x for the journal bearings in a centric position, and as a function of r for the thrust bearing.

rect settings for the various parameters that can affect the stability of a new design — there are 32 of them — has to be made virtually 'in the dark'.

In earlier investigations^[8] a simple stability condition in the form of an analytical expression was derived for a rotor whose bearings are attached to the environment by rigid supports. Since the connection between the spiral-groove bearing via the envelope to the environment is not rigid, we cannot use this simple stability condition.

Fortunately, however, the rotating X-ray-tube anode has a number of symmetries that greatly simplify the problem of stability with flexible bearing supports. It was therefore possible to find an analytical expression for this case as well: a simple relation

for the 'critical speed', the speed at which any increase in speed would lead to the whirling action mentioned above^[9]. The differences between computer calculations and the results of using this simple relation are generally no more than a few per cent. It also turns out that the familiar assumption that the critical speed is twice the lowest resonant frequency of the rotor is by no means true in all cases.

We have used the analytical stability condition that we have derived to test the stability of the anode of the MRC200 tube. The positive results obtained are in complete agreement with practical tests. It will also

be possible to use the method for rapid and early prediction of the likelihood of rotor instabilities in future designs. The physical understanding that results from the relation derived earlier will then indicate methods of avoiding instabilities.

Summary. Spiral-groove bearings can replace the ball bearings in the vacuum of a rotating-anode diagnostic X-ray tube. This increases the maximum load the tube can take and solves a number of other problems, such as limited bearing life, difficulties with anode cooling and noise from the bearings. The spiral-groove bearing has to be lubricated with liquid metal, partly because the vapour pressure of the lubricant should be low, partly because it must conduct the anode current. A gallium alloy has been found to be a successful lubricant. In the new Philips MRC200 diagnostic X-ray tube with spiral-groove bearings more than three times as much energy can be handled during a complete cine investigation as in the SRM 100 tube of the previous generation. This means that an investigation of this type takes only half as long with the new tube, so that there are economic advantages as well as reduced patient stress.

^[8] J. P. Reinhoudt, On the stability of rotor-and-bearing systems and on the calculation of sliding bearings, Thesis, Eindhoven 1972.

^[9] E. A. Muijderman, Algebraic formulas for the threshold and mode of instability and the first critical speed of a simple flexibly supported (overhung) rotor-bearing system, Proc. Conf. on Rotordynamics, Tokyo 1986, 201-210.

Scientific publications

These publications are contributed by staff from the laboratories and other establishments that form part of or are associated with the Philips group of companies. Many of the articles originate from the research laboratories named below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	A
Philips Research Laboratory, Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	B
Philips Natuurkundig Laboratorium, Postbus 80 000, 5600 JA Eindhoven, The Netherlands	E
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	H
Laboratoires d'Electronique Philips, 3 avenue Descartes, 94450 Limeil-Brévannes, France	L
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	N
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	R
Philips Research Laboratories Sunnyvale P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	S

A. H. van Ommen, J. J. M. Ottenheim, A. M. L. Theunissen & A. G. Mouwen	E	Synthesis of heteroepitaxial Si/CoSi ₂ /Si structures by Co implantation into Si	Appl. Phys. Lett. 53	669-671	1988
A. P. M. Kentgens, A. H. Carim & B. Dam	E	Transmission electron microscopy of thin YBa ₂ Cu ₃ O _{7-x} films on (001) SrTiO ₃ prepared by dc triode sputtering	J. Cryst. Growth 91	355-362	1988
A. H. Boonstra & T. N. M. Bernards	E	The dependence of the gelation time on the hydrolysis time in a two-step SiO ₂ sol-gel process	J. Non-Cryst. Solids 105	207-213	1988
A. J. E. M. Janssen	E	On exponentially weighted Toeplitz matrices and their use in the analysis of a frequency-domain, power-compensated adaptive filter	Linear circuits, systems and signal processing: theory and application, C.I. Byrnes <i>et al.</i> (eds), Elsevier Science, Amsterdam	207-216	1988
B. J. van Wees*, L. P. Kouwenhoven*, H. van Houten, C. W. J. Beenakker, J. E. Mooij* (*Univ. of Technol., Delft), C. T. Foxon & J. J. Harris	E,R	Quantized conductance of magnetoelectric subbands in ballistic point contacts	Phys. Rev. B 38	3625-3627	1988
J. C. M. Henning & J. P. M. Ansems	E	Photoionization threshold of the deep donor in Si-doped Al _x Ga _{1-x} As	<i>ibid.</i>	5772-5775	1988
L. W. Molenkamp, G. E. W. Bauer, R. Eppenga & C. T. Foxon	E,R	Exciton binding energy in (Al,Ga)As quantum wells: effects of crystal orientation and envelope-function symmetry	<i>ibid.</i>	6147-6150	1988
A. H. Carim, A. F. de Jong & D. M. de Leeuw	E	Ordering phenomena in the tetragonal superconductor CaBaLaCu ₃ O _{7-δ}	<i>ibid.</i>	7009-7012	1988
A. K. Niessen, A. R. Miedema, F. R. de Boer (Univ. Amsterdam) & R. Boom (Res. Labs, Hoogovens-Groep, IJmuiden)	E	Enthalpies of formation of liquid and solid binary alloys based on 3d metals. IV. Alloys of cobalt	Physica B 151	401-432	1988
G. Keesman	E	Motion estimation based on a motion model incorporating translation, rotation and zoom	Signal processing IV: theories and applications, J.L. Lacoume <i>et al.</i> (eds), Elsevier Science, Amsterdam	31-34	1988
R. N. J. Veldhuis	E	Adaptive restoration of unknown pixels in digital images	Signal processing IV: theories and applications, J.L. Lacoume <i>et al.</i> (eds), Elsevier Science, Amsterdam	1469-1472	1988

Subject index, Volumes 36-44

Figures in bold type indicate the volume number, figures in ordinary type indicate the page number. Subjects dealt with in volumes 1-35 are listed in the indexes included with volumes 10, 15, 20, 25, 30, 35 and 40. The asterisk * indicates that the entry refers to a photograph and caption.

- Absorber plates, black coating of, for solar collectors 43,244
- Abstraction in programming 40,225
- Accordion imager, new solid-state image sensor 43, 1
- Acoustics:**
- vibration patterns and radiation behaviour of loudspeaker cones **36, 1**
 - acoustic surface-wave filters **36, 29**
 - non-rectangular reverberation chamber 37,176
 - increasing reverberation time by multiple-channel amplification **41, 12**
 - noise control in electrical appliances, theory and practice 44,123
 - sound radiation from a vibrating membrane 44,190
- Actuator for anti-lock braking system **36, 74**
- Aircraft, measurement of ozone in **38,131**
- Air pollution, see Environmental science
- Alkali-antimonide films for photocathodes **40, 19**
- Alkali-germanosilicate glass, manufacture of fibres from **36,182**
- Alloys:**
- intermetallic compounds **36,136**
 - heat of formation **36,217**
 - NiFe, for read-out of magnetic tape **37, 42**
 - amalgams for fluorescent lamps **38, 83**
 - tellurium, digital optical recording with **41,313**
 - see also Amorphous alloys
- Alumina, sintered, light transmission of **36, 47**
- Aluminate host lattices for phosphors **37,221**
- Aluminium:**
- electrodeposition of **39, 87**
 - optically smooth, machined on high-precision lathe* **39,183**
 - single-crystal, spark-machined to form test spheres **40,202**
- Aluminium-gallium arsenide:**
- laser for information read-out **39, 37**
 - laser, CQL10, microscope photograph of* **39,324**
 - bistability in quantum-well lasers **44, 76**
 - theory of GaAs/AlGaAs quantum well **44,137**
 - superlattices in short-wavelength lasers **44,268**
 - see also Layered semiconductor structures
- Amalgams for fluorescent lamps **38, 83**
- Amorphous alloys:**
- for magneto-optical recording **42, 37**
 - research on **42, 48**
 - magnetic domains in **44,101**
- Amplitude and phase images, simultaneous, in STEM **37, 1**
- Amplitude modulation (AM) **36,309**
- Analog-to-digital conversion:**
- quantization and coding of analog signals **36,337**
 - delta modulation to PCM converter **37,313**
 - sigma-delta modulation **38,186**
 - digitization of speech **41,201**
 - of hi-fi audio signals, filter for **42,230**
- Anniversary Issue: subjects with a history and a future:**
- special issue { **42,293**
-373
 - introduction **42,293**
 - marginal notes for an anniversary **42,295**
 - television technology from 1936 to 1986 **42,297**
 - history of the Philipsave **42,312**
 - glass — outline of a development **42,316**
 - welding lead-through wires* **42,325**
 - from transistor to IC **42,326**
 - miniaturization of capacitors* **42,335**
 - electric motors in small domestic appliances **42,336**
 - progress in fluorescent lamps **42,342**
 - medical systems in the last half century **42,352**
 - half a century of 'electronification' in telephony systems **42,361**
- Anniversary Issue: 75 years of Philips Research:**
- special issue { **44,237**
-320
 - preface **44,237**
 - miniature beam epitaxy* **44,238**
 - from lamps to ICs **44,239**
 - manufacturing optical fibres by PCVD **44,241**
 - phase-change optical recording **44,250**
 - magnetic fields in medical diagnostics, MR and SQUID **44,259**
 - semiconductor superlattices in short-wavelength lasers **44,268**
 - the eight Laboratories* **44,274**
 - prediction of properties of materials **44,276**
 - sound basis for generation of explanations in expert systems **44,287**
 - advanced HS3 process for fabrication of bipolar ICs **44,296**
 - research on GaAs MESFET circuits at LEP **44,302**
 - power integrated circuits **44,310**
- Anti-lock braking system, fast actuator for **36, 74**
- Arsenic oxides, action in glass fining **40,310**
- Artificial languages in PHLIQA I **38,269**
- ASKA (Automatic System for Kinematic Analysis) used in stress calculations for TV tubes **37, 56**
- ASP, signal processor for digital audio **42,201**
- Aspherics:**
- machining bi-aspheric lenses with COLATH **39,229**
 - preface **41,285**
 - optomechanics for **41,286**
 - design and optical advantages of aspheric surfaces **41,289**
 - fabrication, testing and application of aspheric optical elements **41,296**
- Assembly robot, experimental **40, 33**
- Astronomy, window for ultrasoft X-rays from space **40, 12**
- Audio:**
- annoyance due to modulation noise and drop-outs **37, 29**
 - SPI in FM broadcasting **39,216**
 - manipulation of speech sounds **40,134**
 - Compact Disc Digital Audio **40,149**
 - digital signal processing, computer simulation and listening tests **41, 99**
 - ASP, integrated signal processor for digital audio **42,201**
 - filter for analog-to-digital conversion of hi-fi audio **42,230**
 - see also Acoustics
- AUROS, system for speaker recognition by computer **37,207**
- Automation:**
- flexible { **38,329**
40,237
 - system that learns to recognize two-dimensional shapes **38,356**
 - self-organizing systems **38,364**
 - experimental assembly robot **40, 33**
 - weighing and sorting machine **42,173**
 - automatic segmentation of speech into diphones **43,233**
- Automotive technology:**
- actuator for anti-lock braking system **36, 74**
 - CARIN, car navigation and information system **43,317**
- Avalanche photodiode **36,205**
- Avalanche Triggered Transit diode in TRAPATT oscillator **40, 99**
- Balance, piezoelectric micro- **41,304**
- Barium titanate, PTC effect **38, 73**
- Battery, rechargeable, new type **43, 22**
- Beam manipulation with optical fibres in laser welding **42,262**

- Bearings:**
 spiral-groove, grease-lubricated 39,184
 air, tapered, in metrology . . . 40,338
 measuring radial error of precision air bearings . . . 41,334
 magnetic . . . 41,348
 magnetic, application in Stirling cryogenerators . . . 42, 1
 diagnostic X-ray tube with spiral-groove bearings . . . 44,357
 Bi-asperic lenses, made on COLATH 39,243
 Bioceramic of sintered hydroxylapatite . . . 37,234
 Bipolar IC technology, advanced HS3 process . . . 44,296
 Bismuth-silicon-oxide crystals, large, pulling from the melt . . . 37,250
 Bistability in quantum-well lasers 44, 76
 'Bitter water', making tracks in video tape visible with . . . 40,129
 Black-cobalt coatings for solar collectors . . . 43,244
 Bloch walls in garnet films for fast bubbles . . . 38,211
 BOL, nuclear measurement system at IKO . . . 39,302
Bonding:
 splicing glass fibres for optical communication . . . 38,158
 thermocompression, with IC bonding bit . . . 40,200
 of leadless components to printed boards . . . 40,342
 Boundary-layer capacitors, ceramic 41,338
 Bourdon gauge for hot and corrosive gases . . . 39,344
 Braking system, anti-lock, actuator for . . . 36, 74
 Bridgman anvils for very high pressures . . . 36,251
 Brittle materials, grinding . . . 38,105
 Broadband circulators for VHF and UHF . . . 36,255
 Bubbles, see Magnetic bubbles
 Bucket-brigade memory, see Charge-Transfer Device
 Buildings for submicron IC technology:
 architectural features* . . . 42,172
 technical aspects . . . 42,266
-
- CAD, see Computer-Aided Design**
 Camera tubes, TV, new concept for 39,201
Capacitors:
 ceramic multilayer . . . 41, 89
 switched, in integrated filter for viewdata . . . 41,105
 ceramic boundary-layer . . . 41,338
 miniaturization* . . . 42,335
 Carbon foam . . . 36, 93
 Carbon potentiometers, suspension technology for manufacture of . . . 36,264
 CARIN, car information and navigation system . . . 43,317
 CAROT, digital method of increasing the robustness of TV signals . . . 42,217
 Cars, see Automotive technology
 Cathode-ray tubes for projection television:
 interference filters in . . . 44,201
 phosphor screens in . . . 44,335
 electron guns in . . . 44,348
Cavity resonator:
 for MIP for emission spectroscopy . . . 39, 65
 for microwave measurement of moisture content . . . 40,116
- CCD, see Charge-Transfer Device**
CD-I, see Compact Disc Interactive
Ceramic technology:
 sintered alumina for sodium lamps . . . 36, 47
 manufacture of Ferroxdure . . . 37,157
 sintered hydroxylapatite as bioceramic . . . 37,234
 grinding brittle materials . . . 38,105
 ceramic multilayer capacitors . . . 41, 89
 ceramic boundary-layer capacitors . . . 41,338
 dielectric resonators for microwave integrated oscillators . . . 43, 35
 ceramic differential-pressure transducer . . . 43, 86
Channel multipliers:
 image intensifier for hard X-rays 37,124
 microchannel-plate photomultipliers with subnanosecond characteristics . . . 38,240
Charge-coupled device, see Charge-Transfer Device
Charge-Transfer Device:
 solid-state image sensor . . . 37,303
 P²CCD in storage oscilloscope 40, 55
 accordion imager . . . 43, 1
Chemical analysis, see Chromatography and Inorganic chemical analysis
Chemical modification in surfaces 44, 81
Chemical Vapour Deposition:
 in the manufacture of silica-glass fibres . . . 36,185
 in the manufacture of products of pyrolytic graphite . . . 37,189
 in applying wear-resistant coatings to tool steel . . . 40,204
 growth of silicon films . . . 41, 60
Chromatography:
 for determination of organochlorine residues in milk . . . 36,284
 combination of MIP with gas chromatograph . . . 39, 65
 chromatofocusing for separation of proteins . . . 39,125
 liquid, electrochemiluminescence cells for . . . 40, 69
 Circulators, broadband, for VHF and UHF . . . 36,255
 Clean rooms for LSI* . . . { 37,266
 { 37,272
 Clock radio with LSI* . . . 37,302
 CMOS technology, submicron . . . 42,266
 COLATH, numerically controlled precision lathe . . . 39,229
 Cold cathodes, silicon . . . 43, 49
 Colour television, see Television
 Communication satellites, modulation in . . . 36,359
Compact Disc Digital Audio:
 special issue . . . { 40,149
 { -180
 preface . . . 40,149
 general . . . 40,151
 system aspects and modulation 40,157
 error correction and concealment 40,166
 conversion from digital to analog . . . 40,174
Compact Disc Interactive:
 in CARIN car-navigation system . . . 43,317
 general . . . 44,326
 Compton-scatter radiation, X-ray imaging with . . . 41, 46
Computer-Aided Design:
 with interactive display . . . 36,162
 design of LSI circuits . . . 37,278
 database management for . . . 40,245
 in light optics and electron optics . . . 42, 69
- Computer applications:**
 designing a loudspeaker cone . . . 36, 14
 designing workpieces . . . 36,162
 simulation of control system for water supply . . . 36,273
 optical inspection of connecting-lead patterns for ICs . . . 37, 77
 speaker recognition . . . 37,207
 designing LSI circuits . . . 37,278
 designing TV deflection coils . . . 39,154
 document handling with Megadoc system . . . 39,329
 control of experimental assembly robot . . . 40, 41
 simulation of digital audio signal processing . . . 41, 99
 from powder diagram to structure model on the computer . . . 41,239
 automatic segmentation of speech into diphones . . . 43,233
 expert system for evaluation of infrared spectra . . . 44, 44
 sound radiation from a vibrating membrane . . . 44,190
 see also Image processing and Software
Computers
 modelling and simulation in design of . . . 39,134
 '(4,2)-concept' fault-tolerant computer . . . 41, 1
 Consumer electronics, applied research, the source of innovation in 44,180
 Control, see Measurement and control
 Correction plates for projection TV 39, 15
CQL10:
 information read-out with . . . 39, 37
 microscope photograph* . . . 39,324
 Cryopumps in industrial vacuum technology . . . 39,246
Crystal defects:
 influence on luminescence of GaP 38, 41
 selecting quartz for resonators 43,214
Crystals:
 liquid, for numerical displays . . . 37,131
 bismuth-silicon-oxide, pulling from the melt . . . 37,250
 AlGaAs, for laser . . . 39, 37
 DKDP, for TITUS tube . . . 39, 50
 TTC-cut, for quartz-crystal resonators . . . 40, 1
CTD, see Charge-Transfer Device
 Curvature of reflecting surfaces, measurement of . . . 40,338
CVD, see Chemical Vapour Deposition
-
- Database:**
 computer program for data- { 38,229
 base consultation in English { 38,269
 management for CAD and CAM 40,245
 expert system for evaluation of infrared spectra . . . 44, 44
 Data modem, 'Sematrans' 102* . . . 36,356
Data transmission:
 modulation systems for . . . 36,349
 automatic equalizer for data links . . . 37, 10
 digital modulation stage for . . . 37,291
 baseband, echo canceller for . . . 39,102
 error control in mobile-radio data communication . . . 39,172
 experimental optical-fibre communications network . . . 41,253
 D.C. motors, linear, with permanent magnets . . . 40,329

- Deflection, electro-optic, of laser beam 36,117
- Deflection coils for 30AX colour system 39,154
39,277
- Delta modulation 36,339
37,313
- Diagnostics, see **Medical technology**
- Diamond, spark machining of 40,202
- Diamond die, model of* 40,133
- Differential-pressure transducer, ceramic 43, 86
- Differential pulse-code modulation (DPCM) 36,338
- Diffractionmeter, single-crystal X-ray, PW 1100 38,246
- Digital circuits in video telephone 36,233
- Digital optical recording:**
in Megadoc modular system 39,329
Compact Disc Digital Audio 40,149
with tellurium alloys 41,313
- Digital signal processing:**
digital modulation stage for data transmission 37,291
delta-modulation to PCM converter 37,313
of audio signals: simulation with computer and listening tests 41, 99
special issue on background { 42,101
-144
preface 42,101
growth of a technology 42,103
theoretical background 42,111
special issue on applications { 42,181
-238
preface 42,181
in television receivers 42,183
integrated signal processor for digital audio 42,201
method of increasing robustness of analog TV signal 42,217
filter for analog-to-digital conversion of hi-fi audio 42,230
with PCB 5010 signal processor 44, 1
silicon compiler for designing complex ICs for 44,218
see also **Image processing**
- Digital signals:**
transmission of 36,343
error control in mobile radio 39,172
station and programme identification in FM sound broadcasting 39,216
representation of documents by, with Megadoc 39,329
- Digital systems:**
automatic equalizer for data links 37, 10
echo canceller for data transmission 39,102
equalizer for echo reduction in Teletext 40,319
DIVAC, experimental optical-fibre communications network 41,253
see also **Digital signal processing**
- Digital-to-analog conversion in Compact Disc 40,174
- Digitization of speech 41,201
- Dilution refrigerators for continuous cooling in the millikelvin range 36,104
- Dip soldering electronic components 38,135
- Diphones, automatic segmentation of speech into 43,233
- Dispenser, sodium-vapour 36, 16
- Dispersion measurements on optical fibres for communication 36,211
- Displacement meter based on the laser-Doppler principle 43,180
- Distributed computer systems:**
software aspects 40,262
computations on arrays of processors 40,270
- DIVAC, experimental optical-fibre communication network 41,253
- DKDP crystals for TITUS tube 39, 50
- Document handling with Megadoc system 39,329
- DOD (Droplet On Demand) principle in ink-jet printing 40,192
- Domains:**
observation, in ferroelectrics and ferromagnetics, with SEM 36, 18
optical switching with bismuth-substituted iron garnets 41, 33
magnetic, in amorphous alloys for tape-recorder heads 44,101
see also **Magnetic bubbles**
- Doping by ion implantation 39, 1
- DOR, see **Digital optical recording**
- Double-crucible method for making alkali-germanosilicate glass fibres 36,182
- Double-sideband modulation (DSB) 36,311
- Dye films, organic, for optical recording 41,325
- Echo compensator:**
for data transmission 39,102
for Teletext 40,319
- Echography, medical, transducer for 38,195
- EDDY, computer program for design of electromechanical devices 39, 78
- Electric motors:**
linear d.c. motor with permanent magnets 40,329
linear, in Stirling cryogenerators 42, 1
in small domestic appliances 42,336
- Electrochemistry:**
galvanic effects in wet-chemical etching 38,149
energy production by photoelectrochemical processes 38,160
electrodeposition of aluminium 39, 87
electrochemiluminescence in electrolyte-free solutions 40, 69
electrochemical micromachining 42, 22
new type of rechargeable battery 43, 22
- Electrodeposition of aluminium 39, 87
- Electromechanical devices, design of, by computer 39, 78
- Electromechanical transducers 40,358
- Electron accelerator, linear, at IKO 39,325
- Electron guns:**
in new concept for television camera tubes 39,201
for projection television 44,348
- Electronic components:**
thermal behaviour of, during soldering 38,135
leadless, attachment to printed boards 40,342
- Electronification, half a century of, in telephony systems 42,361
- Electron lithography for LSI 37,334
- Electron microprobe:**
analysis of glass fibres* 40,349
analysis of thin films 42,162
- Electron microscope:**
observation of domains in ferroelectrics and ferromagnetics 36, 18
scanning-transmission 37, 1
CM12/STEM scanning-transmission 43,273
- Electron multipliers, see **Channel multipliers**
- Electron optics:**
pattern generator for LSI 37,334
distortion due to spherical aberration of an electron lens 42, 20
CAD in light optics and electron optics 42, 69
- Electron resists for VLSI 39,346
- Electrophoretic cathode coating 36,264
- Electroradiography, medical 39, 19
- Electrostatic printing 36, 57
- Electrostriction, transducers based on 40,358
- Emission spectrometry by microwave-induced plasma for AES 39, 65
- Energy production by photoelectrochemical processes 38,160
- Engineering technology at IKO 39,315
- Environmental science:**
control of water-purification plant 36,273
measurement of ozone in an aircraft 38,131
- Equalizer, fast automatic, for data links 37, 10
- Erasable magneto-optical recording 42, 37
- Ergonomic lathe* 36,160
- Error correction:**
in mobile-radio data communication 39,172
and concealment in Compact Disc system 40,166
- Etching:**
experimental etching equipment 38, 51
wet-chemical, galvanic effects in 38,149
plasma, in IC technology 38,200
electrochemical micromachining 42, 22
wet-chemical etching of III-V semiconductors 44, 61
- Europe, future industrial development 43, 77
- Evacuated solar collector with heat pipe 40,181
- Expert systems:**
for evaluating infrared spectra 44, 44
sound basis for generation of explanations in 44,287
- Fault-tolerant computer, '(4,2) concept' for 41, 1
- Ferrimagnetic materials:**
garnets for magneto-optic memories 37,197
optical switching with bismuth-substituted iron garnets 41, 33
alloys for erasable magneto-optic recording 42, 37
- Ferroelectrics:**
domains in, observation with SEM 36, 18
ceramic, for analog memory 37, 51
lithium niobate, for holographic information storage 37,109
PTC effect of barium titanate 38, 73
DKDP crystals for TITUS tube 39, 50
for electromechanical transducers with no hysteresis 40,358
- Ferromagnetic materials, domains in, observation with SEM 36, 18
- Ferroxdure 37,157
- Fibre optics, see **Optical fibres**
- Field-effect transistors:**
GaAs, microwave 39,269
monolithic GaAs MESFET circuits 44,302
see also **Layered semiconductor structures**
- Filters:**
based on acoustic surface waves 36, 29
integrated switched-capacitor filter for viewdata 41,105
thin-film reflection filters 41,225
for analog-to-digital conversion of hi-fi audio 42,230
interference, in projection television tubes 44,201

Fining of glass 40,310
 Finite-element method, in calculating stress in TV tubes 37, 56
Fluorescent lamps:
 suspension technology for applying fluorescent coating 36,264
 behaviour of aluminate phosphors for 37,221
 amalgams for 38, 83
 progress in 42,342
 laser diagnostics for low-pressure mercury discharges 43, 62
FM:
 sound broadcasting, station and programme identification in 39,216
 receivers for mono and stereo on single chip 41,169
 Fracture faces of optical glass { 37, 89
 fibres 39,245
 Frequency dividers for UHF 38, 54
 Frequency-division multiplex (FDM) 36,316
 Frequency modulation, see FM
 Frequency stabilization with TTC-cut crystals 40, 1

Gallium arsenide:
 laser for optical communication 36,190
 microwave field-effect transistor 39,269
 bistability in quantum-well lasers 44, 76
 theory of the GaAs/AlGaAs quantum well 44,137
 superlattices in short-wavelength lasers 44,268
 in MESFET ICs 44,302
 see also Layered semiconductor structures
 Gallium arsenophosphide, zinc-diffusion profiles in 37,121
 Gallium nitride, light-emitting diodes based on 37,237
Gallium phosphide:
 zinc-diffusion profiles in 37,121
 influence of crystal defects on luminescence 38, 41
 and GaAs, molecular beam epitaxy of silicon on 43,154
 Galvanic effects in wet-chemical etching 38,149
Garnets:
 ferrimagnetic, for magneto-optic memories 37,197
 garnet films for fast magnetic bubbles 38,211
 optical switching with bismuth-substituted iron garnets 41, 33
 Gas discharge, striations in 44, 89
 Gas-phase epitaxy, see Metal-organic gas-phase epitaxy
 Geiger-Müller counters from IKO 39,296
 Getters, Th-Ce-Al system 36,136
Glass:
 transparent single-point turning 39, 92
 fining 40,310
 mechanism of corrosion by water 42, 59
 outline of a development 42,316
 research on television glass 43,253
 Graphite, pyrolytic, products of 37,189
 Greases for spiral-groove bearings 39,184
 Grinding brittle materials 38,105

HD-MAC, step forward in evolution of television technology 43,197
 Heat of formation of alloys 36,217

Heat pipe:
 in evacuated solar collector 40,181
 in refrigerator-freezer 40,350
 HEIS for determining implantation profiles 39, 1
 High pressures, apparatus for research at 36,245
 High-pressure sodium lamps 39,211
Historical:
 20 years of research on intermetallic compounds 36,136
 IKO, the Institute for Nuclear Physics Research 39,286
 Holst, Gilles, pioneer of industrial research in the Netherlands 40,121
 see also:
 Anniversary Issues
 Philips Technical Review
 50 years ago
 Then and now
Holography:
 holographic display of vibration patterns 36, 1
 lithium niobate for holographic information storage 37,109
 Holst, Gilles, pioneer of industrial research in the Netherlands 40,121
Hydrogen:
 storage in LaNi₅ 36,136
 absorption in intermetallic compounds 36,217
 absorption in rechargeable battery 43, 22
 Hydroxylapatite, sintered, as bioceramic 37,234

IC, see Integrated circuits
 I²L, see Injection logic
Image intensifier:
 for hard X-rays 37,124
 X-ray, with large input format 41,137
Image processing:
 camera for, with edge-enhancement 37,180
 special issue { 38,289
 -371
 preface 38,289
 a challenge 38,291
 digital image enhancement 38,298
 Dot Scan CCTV for, in real time 38,310
 digital, universal instrument for 38,326
 experiment in flexible automation 38,329
 flashing tomosynthesis 38,338
 computerized mammogram processing 38,347
 system that learns to recognize two-dimensional shapes 38,356
 self-organizing systems 38,364
 dual-energy X-ray diagnostics 42,274
 interactive MR image synthesis 43, 95
 Image projector, electron, for LSI 37,347
Image sensors:
 solid-state, with resistive electrodes 37,303
 accordion imager 43, 1
 mobile system for image bulk storage 43,260
 Implantation, see Ion implantation
 INDA, software for flexible automation 40,237
 Indentation measurements on thin films 42, 85
 Index profiling of optical fibres 36,211
 Industrial development of Europe 43, 77

Infrared:
 thermography 37,241
 thin-film reflection filters 41,225
 expert system for evaluation of infrared spectra 44, 44
Injection logic:
 CAD system for I² L circuits. 37,290
 I² L circuit for digital modulation stage 37,291
 Injection-moulding process, analysis 44,212
 Ink-jet printing 40,192
Inorganic chemical analysis:
 measurement of ozone in an aircraft 38,131
 emission spectrometry with MIP 39, 65
 high-resolution X-ray diffractometer 41,183
 electron-probe microanalysis of thin films 42,162
Institute for Nuclear Physics Research (IKO):
 special issue { 39,285
 -328
 preface 39,285
 historical 39,286
 cyclotron 39,290
 radioisotopes 39,294
 Geiger-Müller counters 39,296
 semiconductor detectors 39,298
 BOL 39,302
 electronics for nuclear physics 39,312
 engineering technology 39,315
 ion-beam technology 39,320
 linear electron accelerators 39,325
Integrated circuits:
 pnpn elements for telephone switching 36,291
 optical inspection of connecting-lead patterns 37, 77
 in PM 2517 digital multimeter 38,181
 plasma etching in IC technology 38,200
 echo canceller for data transmission 39,102
 equalizer for echo reduction in Teletext 40,319
 integrated switched-capacitor filter for viewdata 41,105
 FM receivers on single chip 41,169
 digitization of speech 41,201
 expansion of IC activities at Philips Research Laboratories* 42,172
 signal processor for digital audio 42,201
 filter for analog-to-digital conversion of hi-fi audio 42,230
 from transistor to IC 42,326
 dielectric resonators for microwave oscillators 43, 35
 digital signal processors 44, 1
 complex, design of for digital signal processing 44,218
 advanced HS3 bipolar-IC technology 44,296
 GaAs MESFET ICs 44,302
 power integrated circuits 44,310
 see also:
 LSI
 Submicron IC technology
 VLSI
 Interactive display in computer-aided design 36,162
 Interconnection patterns for ICs, optical inspection of 37, 77
 Interfaces, looking at 43,220
 Interference filters in projection-television tubes 44,201
Intermetallic compounds:
 20 years of research 36,136
 computer calculation of crystal structure from powder diagram 41,239
 Ion-beam mixing 44, 24
 Ion-beam system at FOM* 39,319

- Ion-beam technology at IKO . . . 39,320
- Ion implantation:
in semiconductors . . . 39, 1
an open 800-kV machine . . . 43,169
- Ionography, with compressed gas,
system for medical electroradiog-
raphy . . . 39, 19
- Iridescence in technology and nature* 41,149
- Iron garnets, bismuth-substituted,
optical switching with . . . 41, 33
- Isotopes for determining zinc-diffu-
sion profiles in GaP and GaAsP . . . 37,121
- Lacquers:**
photopolymerizable, for Laser-
Vision discs . . . 40,298
X-ray resists for VLSI . . . 41,150
for photolithography, polymer
chemistry of . . . 42,149
- LAN, Local-Area Network . . . 43, 10
- Langmuir trough for building mono-
molecular layers . . . 36, 44
- Large-screen projector with laser beam 36,117
- Lasers:**
electro-optic deflection of laser
beam . . . 36,117
beam manipulation with optical
fibres in laser welding . . . 42,262
laser diagnostics for low-pressure
mercury discharges . . . 43, 62
laser-Doppler displacement meter 43,180
module for 4-Gbit/s optical
communications . . . 44,162
see also Semiconductor lasers
- LaserVision:**
scanner for player* . . . 37, 90
disc manufacture . . . 40,287
photopolymerizable lacquers for
discs . . . 40,298
- Lathe:**
ergonomic* . . . 36,160
high-precision, COLATH . . . 39,229
fabrication of aspheric optical
elements with COLATH . . . 41,296
- Layered semiconductor structures:**
special issue . . . { 43,109
-165
preface . . . 43,109
research on . . . 43,111
metal-organic vapour-phase epi-
taxy (MO-VPE) of III-V semi-
conductors . . . 43,118
MO-VPE with novel reactor and
characterization . . . 43,133
molecular beam epitaxy (MBE)
of GaAs and AlGaAs . . . 43,143
MBE of Si on GaP and GaAs . . . 43,154
- Leaching of glass, mechanism for . . . 42, 59
- Lead-through wires, welding of* . . . 42,325
- Lectures:**
H.B.G. Casimir . . . 40,121
W. Dekker . . . 43, 77
J. F. Dijkman . . . 44,212
Bernard Dixon . . . 38, 17
S. van Houten . . . 44,180
G. Lorenz . . . 44, 16
W. Martienssen . . . 38, 25
F. Meijer . . . 43,220
A. E. Pannenberg . . . 38, 33
G. E. Thomas . . . 44, 51
- LED, see Light-emitting diodes
- Lens surfaces, inspecting shape of* 41,224
- Levitation in magnetic bearings . . . 41,348
- Light beam, deflection of, with Kerr
cells . . . 36,117
- Light-emitting diodes:**
for optical communication . . . 36,190
of GaN . . . 37,237
- Light guides in process control . . . 43, 58
- Light modulation in optical-fibre
transmission system . . . 36,201
- Light optics and electron optics,
CAD in . . . 42, 69
- Light transmission of sintered
alumina . . . 36, 47
- Liquid crystals for numerical displays 37,131
- Listening room for appraisal of digital
audio signal processing . . . 41, 99
- Lithium niobate for holographic
information storage . . . 37,109
- Lithography:**
methods for IC production . . . 37,270
Silicon Repeater for LSI . . . 37,330
electron-beam pattern generator 37,334
electron-image projector . . . 37,347
X-ray, for VLSI . . . 41,150
optics of Silicon Repeater . . . 41,268
polymer chemistry of lacquers
for photolithography . . . 42,149
- Local-area network, PHILAN, based
on fibre-optic ring . . . 43, 10
- Logic analyser, PM 3543* . . . 40,286
- Logic circuits, see Injection logic
- Loudspeaker cones, vibration pat-
terns and radiation behaviour of . . . 36, 1
- LSI:**
special issue . . . { 37,265
-356
preface . . . 37,265
revolution in electronics . . . 37,267
abbreviations and acronyms . . . 37,277
design by CAD . . . 37,278
digital modulation stage for data
transmission . . . 37,291
image sensor with resistive elec-
trodes . . . 37,303
delta modulation/PCM con-
verter . . . 37,313
Silicon Repeater . . . 37,330
electron-beam pattern generator 37,334
electron-image projector . . . 37,347
- Luminescence:**
of GaP, influence of crystal de-
fects on . . . 38, 41
electrochemi-, in electrolyte-free
solutions . . . 40, 69
see also Phosphors
- Machine-print characters, image pro-
cessing of, with OCR camera . . . 37,180**
- Machining, spark . . . 40,199
- Machining, transparent, of glass . . . 39, 92
- Magnetic bubbles**
single-mask bubble memory with
rotating field control . . . 36,149
generation and manipulation by
microwaves . . . 37, 38
fast, garnet films for . . . 38,211
- Magnetic domains, see Domains
- Magnetic fields of TV deflection coils,
measurement of . . . 39,277
- Magnetic heads:**
manufacture on laboratory scale 44,151
multi-track, in thin-film techno-
logy . . . 44,169
- Magnetic properties of microscopic
particle, measurement of . . . 39, 48
- Magnetic recording:**
of video signals . . . 36,326
annoyance due to modulation
noise and drop-outs in sound
recording . . . 37, 29
read-out of magnetic tape by
magnetoresistance effect . . . 37, 42
making video tracks visible . . . 40,129
- magnetic domains in amorphous
alloys for tape-recorder heads 44,101
- manufacture of magnetic heads
on laboratory scale . . . 44,151
- multi-track magnetic heads in
thin-film technology . . . 44,169
- Magnetic resonance:**
proton NMR tomography . . . 41, 73
interactive image synthesis . . . 43, 95
magnetic fields in medical diag-
nostics . . . 44,259
- Magnetic tape:**
production of . . . 36,268
mercury porosimetry of . . . 41,260
- Magneto-optic storage wafer MOPS 37,197
- Magneto-optical recording . . . 42, 37
- Magnetoresistance effect, read-out of
magnetic tape by . . . 37, 42
- Mammograms, computerized proces-
sing of . . . 38,347
- Manometers, see Pressure measure-
ments
- Masks for integrated circuits:**
for fabrication of digital mod-
ulation stage . . . 37,300
electron-beam pattern generator
for LSI . . . 37,334
removal of, by plasma etching 38,200
checking computer design of* 38,290
X-ray lithography for VLSI . . . 41,150
- Materials testing:**
metal/ceramic X-ray tubes for . . . 41, 24
X-ray imaging for, with Com-
pton-scatter radiation . . . 41, 46
determining Von Mises stress
from neck of a test specimen . . . 42, 11
quantitative measurements by
the Schlieren method . . . 43,184
selecting quartz for resonators 43,214
predicting the properties of mat-
erials . . . 44,276
- MCR, increasing reverberation time
by multiple-channel amplification 41, 12
- Measurement and control:**
control of water-purification
plant . . . 36,273
control of experimental robot . . . 40, 41
process control with light
guides . . . 43, 58
ceramic differential-pressure
transducer . . . 43, 86
- Mechanization, see Automation**
- Medical technology:**
transducer for medical echo-
graphy . . . 38,195
digital enhancement of X-ray
pictures . . . 38,298
flashing tomosynthesis of X-ray
pictures . . . 38,338
mammogram processing . . . 38,347
electroradiography . . . 39, 19
TOMOSCAN 310* . . . 40,253
X-ray imaging with Compton-
scatter radiation . . . 41, 46
proton NMR tomography . . . 41, 73
metal/ceramic diagnostic X-ray
tube . . . 41,126
X-ray image intensifier with large
input format . . . 41,137
dual-energy X-ray diagnostics 42,274
medical systems in the last half-
century . . . 42,352
interactive MR image synthesis 43, 95
magnetic fields in medical diag-
nostics: MR and SQUID . . . 44,259
diagnostic X-ray tube with spiral-
groove bearings . . . 44,357
- Megadoc, modular system for elec-
tronic document handling . . . 39,329
- Membrane, vibrating, sound radia-
tion from . . . 44,190

- Memories:**
 single-mask bubble memory with rotating-field control . . . 36,149
 analog, ferroelectric . . . 37, 51
 lithium niobate, for holographic information storage . . . 37,109
 magneto-optic, of discrete-bit type . . . 37,197
 DOR disc in Megadoc modular system . . . 39,329
 storage oscilloscope with P² CCD . . . 40, 55
 DOR with tellurium alloys . . . 41,313
 submicron IC technology for development of 1-Mbit SRAM . . . 42,266
 256-kbit SRAM . . . 44, 33
- Mercury discharges, low-pressure, laser diagnostics for . . . 43, 62
- Mercury lamps, see **Fluorescent lamps**
- Mercury porosimetry of magnetic tape . . . 41,260
- MESFET circuits, GaAs, research on at LEP . . . 44,302
- Metal/ceramic X-ray tubes:**
 for non-destructive testing . . . 41, 24
 for medical diagnostics . . . 41,126
- Metallurgical quantities, values of . . . 38,257
- Metal-organic vapour-phase epitaxy:**
 of multilayer structures with III-V semiconductors . . . 43,118
 with novel reactor and characterization of multilayer structures . . . 43,133
- Metals:**
 20 years of research on inter-metallic compounds . . . 36,136
 heat of formation of alloys . . . 36,217
 interconnection in ICs . . . 37,267
 atom as building block . . . 38,257
 from powder diagram to structure model on the computer . . . 41,239
 research on amorphous alloys . . . 42, 48
- Metastable phases . . . 43,304
- Microbalance, piezoelectric . . . 41,304
- Microchannel plate in subnanosecond photomultiplier tubes . . . 38,240
- Microcomputer Development System PM 4421* . . . 40,269
- Microcomputer for intercom system* . . . 37,275
- Micromachining, electrochemical . . . 42, 22
- Microwaves:**
 circulators, broadband . . . 36,255
 modulation in microwave links . . . 36,357
 generation and manipulation of magnetic domains . . . 37, 38
 microwave-induced plasma for emission spectrometry . . . 39, 65
 GaAs microwave FET . . . 39,269
 TRAPATT oscillator . . . 40, 99
 measurement of moisture content in process materials . . . 40,112
 dielectric resonators for integrated microwave oscillators . . . 43, 35
- Milk, determination of organochlorine residues in . . . 36,284
- MIP (Microwave-Induced Plasma) for emission spectrometry . . . 39, 65
- Mobile image-storage system . . . 43,260
- Mobile-radio data communication, error control in . . . 39,172
- Modulation:**
 light, in optical-fibre transmission systems . . . 36,201
 special issue . . . { 36,305
 -362
 preface . . . 36,305
 introduction . . . 36,307
 of sinusoidal carrier . . . 36,309
 of pulse trains . . . 36,329
 quantization and coding of analog signals . . . 36,337
 transmission of digital signals . . . 36,343
- in telecommunication . . . 36,353
 in data transmission . . . 37, 13
 digital modulation stage for data transmission . . . 37,291
 delta modulation/PCM converter with LSI . . . 37,313
 in Compact Disc . . . 40,157
- Modulation noise in magnetic sound recording . . . 37, 29
- Moisture content of process materials, measured with microwaves . . . 40,112
- Molecular beam epitaxy (MBE)**
 of multilayer structures with GaAs and AlGaAs . . . 43,143
 of silicon on GaP and GaAs . . . 43,154
 silicon-MBE equipment* . . . 44,238
- Monomolecular layers, Langmuir trough for . . . 36, 44
- MOPS, magneto-optic storage wafer . . . 37,197
- MOS devices, behaviour of oxide film . . . 43,330
- MR, see **Magnetic resonance**
- MRH (Magneto-Resistive Head) for read-out of magnetic tape . . . 37, 42
- Multimeter, PM2517 automatic digital . . . 38,181
- Multiplexers and demultiplexers in optical communication systems . . . 43,344
- Multi-track magnetic heads . . . 44,169
-
- Navigation and information system for cars, CARIN . . . 43,317
- Network structure of polyepoxides . . . 44,110
- Nickel-hydride cell, new type of rechargeable battery . . . 43, 22
- NMR, see **Magnetic resonance**
- Noise:**
 annoyance due to modulation noise in magnetic sound recording . . . 37, 29
 due to optical feedback in semiconductor lasers . . . 43,292
- Nuclear-physics electronics at IKO . . . 39,312
- Nucleation and growth of silicon films by CVD . . . 41, 60
- Numerical display with liquid crystals . . . 37,131
-
- O-BUS, system for flexible public transport by on-call buses . . . 40,231
- OCR (Optical Character Recognition), edge-enhancing double-focus camera . . . 37,180
- Optical communication via glass fibres:**
 special issue . . . { 36,177
 -216
 preface . . . 36,177
 general . . . 36,178
 manufacture of glass optical fibres . . . 36,182
 semiconductor lasers for . . . 36,190
 light modulation and injection . . . 36,201
 avalanche photodiode as detector . . . 36,205
 testing optical fibres by dispersion measurements . . . 36,211
 components for glass optical-fibre circuits . . . 40, 46
 DIVAC experimental network . . . 41,253
 integrating sphere for measuring losses* . . . 41,347
 system with wavelength-division multiplex and minimized insertion loss . . . { 42,245
 43,344
- transmission in the 1.55- μ m window in single-mode fibre* . . . 42,286
- PHILAN, local-area network based on fibre-optic ring . . . 43, 10
 up to 4-Gbit/s, laser module for . . . 44,162
- Optical feedback, noise due to, in semiconductor lasers . . . 43,292
- Optical fibres:**
 obtaining smooth fracture faces . . . 37, 89
 machine for hot splicing . . . 38,158
 precision fracture of* . . . 39,245
 analysis of, with electron microscope* . . . 40,349
 device for stripping protective coatings from . . . 41,124
 polymer chemistry of protective coatings . . . 42,149
 beam manipulation in laser welding . . . 42,262
 application of light guides in process control . . . 43, 58
 manufacture by PCVD process . . . 44,241
 see also **Optical communication via glass fibres**
- Optical inspection of connecting-lead patterns for ICs . . . 37, 77
- Optical integrating sphere for measuring losses in optical fibres* . . . 41,347
- Optical recording:**
 manufacture of LaserVision discs . . . 40,287
 lacquers for LaserVision discs . . . 40,298
 organic-dye films for optical recording . . . 41,325
 magneto-optical, erasable . . . 42, 37
 future trends . . . 44, 51
 by crystalline/amorphous phase change . . . 44,250
 see also **Digital optical recording**
- Optical switching with bismuth-substituted iron garnets . . . 41, 33
- Optical technology:**
 turning glass for aspheric lenses . . . 39, 92
 making bi-aspheric lenses with COLATH . . . 39,243
 for fabrication of aspheric optical elements . . . 41,286
- Organic-dye films for optical recording . . . 41,325
- Orthogonal transformation of TV pictures in real time . . . 38,119
- Oscillators, integrated microwave, dielectric resonators for . . . 43, 35
- Oscilloscope, digital storage, with P²CCD . . . 40, 55
- Oxide film in MOS devices . . . 43,330
- Ozone, measurement of, in an aircraft . . . 38,131
-
- Parallel computer programs . . . { 40,254
 40,278
- Passivation layers, measurement of stress in . . . 39,130
- Pattern generator, electron-beam, for LSI . . . 37,334
- P²CCD, in storage oscilloscope . . . 40, 55
- PCVD for manufacture of optical fibres . . . 44,241
- Perception:**
 measurement of visual conspicuity . . . 36, 71
 OCR camera with edge-enhancement for image processing . . . 37,180
 speaker recognition by computer . . . 37,207
 manipulation of speech sounds . . . 40,134
- Permanent magnets:**
 intermetallic compounds for . . . 36,136
 Ferroxdure for . . . 37,157
 linear d.c. motor with . . . 40,329

- Personal computer, P 2000* . . . 40,261
- Pesticides, milk monitor for . . . 36,284
- Phase and amplitude contrast, simultaneous, in STEM . . . 37, 1
- Phase modulation . . . 36,318
- Phases, metastable, and thermodynamic equilibrium . . . 43,304
- Phenol synthesis and photomorphogenesis . . . 38, 89
- PHIDAS, database management system for CAD/CAM . . . 40,245
- PHIDIAS, software aspects of . . . 40,262
- PHILAN, a fibre-optic local-area network . . . 43, 10
- Philips Technical Review ceases publication . . . 44,325
- Philips Technical Review 50 years ago:
- introduction . . . 42,160
 - experimental television transmitter and receiver* . . . 42,161
 - loudspeaker and sound-amplifying installation on 'Normandie'* . . . 42,161
 - optical models for studying acoustics of theatres* . . . 42,265
 - practical applications of X-rays for the examination of materials* . . . 42,265
 - demonstration of superheterodyne reception* . . . 43, 9
- Philishave, history of . . . 42,312
- PHLIQA 1:**
- organization and performance . . . 38,229
 - artificial languages and translation operations . . . 38,269
- Phosphors:**
- aluminate, for fluorescent lamps . . . 37,221
 - pigmentation of, for colour TV . . . 40, 48
 - for projection television . . . 44,335
- Photocathodes:**
- caesium iodide, for electron-image projector . . . 37,348
 - alkali-antimonide films for . . . 40, 19
 - X-ray image intensifier with large input format . . . 41,137
- Photodiodes, avalanche . . . 36,205
- Photoelectrochemical energy production . . . 38,160
- Photoemission of alkali-antimonide films . . . 40, 19
- Photomasks, see **Masks for integrated circuits**
- Photomorphogenesis . . . 38, 89
- Photomultipliers, ultra-fast, with microchannel plate . . . 38,240
- Photopolymerization:**
- in manufacture of LaserVision discs . . . 40,287
 - photopolymerizable lacquers for LaserVision discs . . . 40,298
- Picture processing, see **Image processing**
- Picture tubes, see **Television**
- Piezoelectric microbalance . . . 41,304
- Piezoelectric weighing and sorting of small spherical particles . . . 42,173
- Pigmentation of TV phosphors . . . 40, 48
- Plants, phenol synthesis and photomorphogenesis in . . . 38, 89
- Plasma:**
- plasma method for making glass optical fibres . . . 36,182
 - plasma etching in IC technology . . . 38,200
 - microwave-induced, for emission spectrometry . . . 39, 65
 - PCVD process for manufacture of optical fibres . . . 44,241
- Polyepoxides . . . 44,110
- Polymer chemistry in the electrical industry . . . 42,149
- Porosimeter measurements on magnetic tape . . . 41,260
- Powder diagram, calculating crystal structure by computer from . . . 41,239
- Power ICs . . . 44,310
- Pressure measurements:**
- fast gauge for hot and corrosive gases . . . 39,344
 - ceramic differential-pressure transducer . . . 43, 86
- Pressure vessels for solid-state physics research . . . 36,245
- Priinting:**
- electrostatic . . . 36, 57
 - ink-jet . . . 40,192
- Process control with light guides . . . 43, 58
- Projection television:**
- correction plates for . . . 39, 15
 - interference filters in projection tubes . . . 44,201
 - phosphor screens in projection tubes . . . 44,335
 - electron guns in projection tubes . . . 44,348
- Protective coatings for optical fibres:**
- removal . . . 41,124
 - polymer chemistry . . . 42,149
- Proteins, separation of, by chromatofocusing . . . 39,125
- Proton NMR tomography . . . 41, 73
- PSD (Picture Store and Display) instrument for digital image processing . . . 38,326
- PTC effect of barium titanate . . . 38, 73
- Pulse-amplitude modulation (PAM) . . . 36,330
- Pulse-code modulation (PCM) . . . 36,337
- Pulse-frequency and pulse-duration modulation (PFM and PDM) . . . 36,332
- Purification plant for water, control of . . . 36,273
- PVD (Physical Vapour Deposition) of low-friction wear-resistant films . . . 41,186
- PW1100, single-crystal X-ray diffractometer . . . 38,246
- Pyrolytic graphite, products of . . . 37,189
- Quantum well:**
- lasers, bistability in . . . 44, 76
 - GaAs/AlGaAs, theory of . . . 44,137
 - see also **Layered semiconductor structures**
- Quartz, selection of, for resonators . . . 43,214
- Quartz-crystal resonators with TTC-cut crystals . . . 40, 1
- Quartz glass optical fibre, manufacture . . . 36,182
- Question-answering system for database consultation in English . . . 38,229
- Radar, 'Signal' Automatic Radar Processing system*** . . . 40,218
- Radial error of precision air bearings, measurement of . . . 41,334
- Radio:**
- station and programme identification in FM sound broadcasting . . . 39,216
 - FM receivers for mono and stereo on single chip . . . 41,169
- Radioisotopes from IKO . . . 39,294
- Recognition:**
- of speakers by computer . . . 37,207
 - of two-dimensional shapes, by learning system . . . 38,356
 - picture, with self-organizing systems . . . 38,364
 - expert system for evaluation of infrared spectra . . . 44, 44
- Reflection filters, thin-film . . . 41,225
- Reflection reduction on television screens* . . . 42, 58
- Refrigeration technology:**
- continuous cooling in the millikelvin range . . . 36,104
 - critical review . . . 37, 91
 - refrigerator-freezer with heat pipe . . . 40,350
 - Stirling cryogenerators with linear drive . . . 42, 1
- Repeater projector, see **Silicon Repeater**
- Resistive electrodes for image sensor . . . 37,303
- Resists, electron, for VLSI . . . 39,346
- Resonators:**
- with TTC-cut quartz crystals . . . 40, 1
 - dielectric, for integrated microwave oscillators . . . 43, 35
 - selecting quartz for . . . 43,214
- Reverberation chamber, non-rectangular . . . 37,176
- Reverberation time, increased by multiple-channel amplification . . . 41, 12
- Robot, experimental, for assembly . . . 40, 33
- Rotating-field control, single-mask bubble memory with . . . 36,149
- Roughness measurements by Schlieren method . . . 43,184
- SAMEN/SAMO** . . . 39,134
- Satellite television, 12-GHz receivers for . . . 39,257
- Scandium for X-ray tube anode* . . . 41,267
- Scanning-transmission electron microscope:**
- simultaneous phase and amplitude images in . . . 37, 1
 - CM12/STEM . . . 43,292
- Schlieren method, quantitative measurements with . . . 43,184
- Semtrans 102* . . . 36,356
- Semiconductor lasers:**
- for optical communication . . . 36,190
 - coupling to optical fibre . . . 36,201
 - for information read-out . . . 39, 37
 - microscope photograph of CQL10* . . . 39,324
 - noise due to optical feedback in semiconductor lasers . . . 43,292
 - for visible light* . . . 44, 23
 - quantum-well, bistability in . . . 44, 76
 - short-wavelength, superlattices in . . . 44,268
 - see also:
 - LaserVision**
 - Layered semiconductor structures**
 - Optical recording**
- Seventy-five years of Philips Research, see **Anniversary Issue**
- Shape inspection of lens surfaces* . . . 41,224
- Shape memories of NiTe . . . 36,136
- Shavers, history of the Philishave . . . 42,312
- Signal processors:**
- ASP, for digital audio . . . 42,201
 - digital, PCB 5010 . . . 44, 1
- Silica-glass optical fibre, manufacture . . . 36,185

- Silicon:**
 avalanche photodiode of, for optical communication . . . 36,205
 in TRAPATT oscillator . . . 40, 99
 films, nucleation and growth of, by CVD . . . 41, 60
 cold cathodes . . . 43, 49
 molecular beam epitaxy of, on GaP and GaAs . . . 43,154
 see also **Integrated circuits**
 Silicon compiler for design of complex ICs . . . 44,218
Silicon Repeater:
 for LSI . . . 37,330
 optical aspects of . . . 41,268
SIMS for determining doping profiles . . . 39, 1
 Single-mask bubble memory with rotating-field control . . . 36,149
 Single-mode optical fibre, transmission in the 1.55- μ m window* . . . 42,286
 Single-sideband modulation (SSB) . . . 36,313
 Sintering, see **Ceramic technology**
 Skull-melting, method for continuous melting of glass . . . 42, 93
 Sodium lamps, high-pressure . . . 39,211
 Sodium-vapour dispenser . . . 36, 16
Software:
 ASKA, for stress calculations on TV tubes . . . 37, 56
 for database consultation in English . . . 38,229
 EDDY, for design of electromagnetic devices . . . 39, 78
 SAMEN/SAMO, for simulating computer systems . . . 39,134
 for simulating telephone cables . . . 40, 85
 special issue . . . 40,217
 -286
 preface . . . 40,217
 general . . . 40,219
 abstraction . . . 40,225
 O-BUS, system for public transport with on-call buses . . . 40,231
 INDA, software tool for production engineer . . . 40,237
 database management system for CAD and CAM . . . 40,245
 parallel programs . . . 40,254
 software aspects of PHIDIAS system . . . 40,262
 PM4421, aid in microcomputer development* . . . 40,269
 distributed computations on arrays of processors . . . 40,270
 transformation methods for improving parallel programs . . . 40,278
 for calculating crystal structure from powder diagram . . . 41,239
 expert system for evaluation of infrared spectra . . . 44, 44
 Solar cells, regenerative . . . 38,166
Solar collectors:
 evacuated, with heat pipe . . . 40,181
 black-cobalt coatings for . . . 43,244
Soldering:
 thermal behaviour of electronic components during soldering . . . 38,135
 of leadless components to printed boards . . . 40,342
SON lamps (high-pressure sodium lamps) . . . 39,211
 Sound, see **Acoustics and Audio**
 Space-division multiplex (SDM) system for telephone switching . . . 36,294
 Spark machining . . . 40,199
 SPARX system, speech studies with . . . 40,134
 Speaker recognition by computer . . . 37,207
Spectrometry:
 HEIS and SIMS for determining doping profiles . . . 39, 1
 by microwave-induced plasma for atomic emission . . . 39, 65
 Raman, for investigating glass fining . . . 40,310
 scandium for X-ray tube anode for X-ray fluorescence* . . . 41,267
Speech:
 manipulation of speech sounds . . . 40,134
 digitization of . . . 41,201
 automatic segmentation of, into diphones . . . 43,233
 Spherical aberration of an electron lens, distortion due to . . . 42, 20
 Spherical particles, weighing and sorting . . . 42,173
 SPI in FM sound broadcasting . . . 39,216
 Splicing, see **Bonding**
 SQUID (Superconducting Quantum Interference Device) . . . 44,259
SRAM, Static Random-Access Memory:
 256-kbit SRAM . . . 44, 33
 in submicron IC technology* . . . 44,150
 Station and programme identification in FM sound broadcasting . . . 39,216
STEM, see Scanning-transmission electron microscope
 Strain gauges, thin-film . . . 39, 94
 Striations in a gas discharge . . . 44, 89
 Strontium titanate in ceramic boundary-layer capacitors . . . 41,338
Submicron IC technology:
 new centre for . . . 42,266
 the 256-kbit SRAM, important step on the way to . . . 44, 33
 SRAM in* . . . 44,150
Superconductivity:
 in intermetallic compounds . . . 36,136
 and superfluidity . . . 37, 91
 Superlattices in short-wavelength semiconductor lasers . . . 44,268
Surfaces:
 aluminium, optically smooth* . . . 39,183
 reflecting, curvature of . . . 40,338
 aspheric, design and optical advantages of . . . 41,289
 mechanism of the corrosion of glass by water . . . 42, 59
 quantitative measurements by the Schlieren method . . . 43,184
 looking at interfaces . . . 43,220
 chemical modification in . . . 44, 81
 Surface tension of metals . . . 38,257
 Surface waves, acoustic, in filters . . . 36, 29
 Suspension technology . . . 36,264
 { 39,291
 Synchrocyclotron at IKO . . . 39,308
- Teletext:**
 LSI decoder for reception of* . . . 37,312
 automatic equalizer for echo reduction . . . 40,319
Television:
 stress calculations for picture tubes . . . 37, 56
 solid-state image sensor . . . 37,303
 real-time orthogonal transformation of colour-TV pictures . . . 38,119
 Dot Scan CCTV for real-time image processing . . . 38,310
 deflection coils of 30AX colour-picture system . . . 39,154
 new concept for camera tubes . . . 39,201
 12-GHz receivers for satellite TV . . . 39,257
 measuring magnetic fields of TV deflection coils . . . 39,277
 games with computer* . . . 40,230
 reduction of screen reflections* . . . 42, 58
 receivers, digital signal processing in . . . 42,183
 digital method for obtaining robust television signals . . . 42,217
 technology, from 1936 to 1986 . . . 42,297
 accordion imager . . . 43, 1
 test decor* . . . 43,192
 HD-MAC . . . 43,197
 research on television glass . . . 43,253
 see also **Projection television**
 Tellurium alloys for digital optical recording . . . 41,313
 Tensile test, determining Von Mises stress from neck of specimen . . . 42, 11
Then and now (1937-1987):
 lighting the Eiffel tower* . . . 43, 61
 high-pressure mercury lamps* . . . 43, 94
 measuring instruments* . . . 43,183
 television projection tubes* . . . 43,213
 transmitting valves* . . . 43,243
 permanent magnets* . . . 43,303
 high-voltage rectifiers* . . . 43,343
Then and now (1938-1988):
 television cameras* . . . 44, 15
 plastics in vacuum cleaners* . . . 44, 43
 lamps for phototherapy* . . . 44, 75
 antennas* . . . 44,122
 assembly of thermionic devices* . . . 44,161
 car radios* . . . 44,179
Then and now (1939-1989):
 Philishave* . . . 44,211
 television receivers* . . . 44,334
Theoretical physics:
 GaAs/AlGaAs quantum well . . . 44,137
 prediction of properties of materials . . . 44,276
 Thermodynamic equilibrium and metastable phases . . . 43,304
 Thermography, macro- and micro- . . . 37,241
Thin films and coatings:
 strain gauges . . . 39, 94
 CVD for applying wear-resistant coatings to tool steel . . . 40,204
 nucleation and growth of silicon films by CVD . . . 41, 60
 applying low-friction wear-resistant films by PVD . . . 41,186
 thin-film reflection filters . . . 41,225
 tellurium alloys for digital optical recording . . . 41,313
 organic-dye films for optical recording . . . 41,325
 indentation measurements on . . . 42, 85
 electron-microprobe analysis of oxide film in MOS devices . . . 43,330
 improved adhesion of lubricating films with ion-beam mixing . . . 44, 24
 multi-track magnetic heads . . . 44,169
 see also **Layered semiconductor structures**
- Telecommunication:**
 broadband circulators for VHF and UHF . . . 36,255
 modulation in . . . 36,353
 advanced, technological aspects . . . 44, 16
 see also **Data transmission and Telephony**
Telephony:
 switching telephone signals . . . 36,291
 delta modulation/PCM converter with LSI . . . 37,313
 computer-aided research on multiwire cables . . . 40, 85
 digital exchange* . . . 40,224
 half a century of 'electronification' in telephony systems . . . 42,361
 see also **Optical communication via glass fibres and Video telephony**

Time-division multiplex (TDM) with pulse-amplitude modulation . . . 36,330
 Titanium carbide and nitride, wear-resistant coatings on tool steel . . . 40,204
 TITUS tube, DKDP crystals for . . . 39, 50
 Tomography, NMR . . . 41, 73
 Tomosynthesis, flashing . . . 38,338
 Tool steel, application of wear-resistant coatings to . . . 40,204
Transducers:
 for medical echography . . . 38,195
 with thin-film strain gauges . . . 39, 94
 electromechanical, with no hysteresis . . . 40,358
 Transformation methods for improving parallel computer programs . . . 40,278
 Transistor technology . . . 42,326
 Transition metals, alloying behaviour of . . . 36,217
 Translation operations in PHLIQA I . . . 38,269
 Transmitting-valve grids of pyrolytic graphite . . . 37,189
 Transport, flexible public, O-BUS system with on-call buses . . . 40,231
 Transport reactions, chemical, in incandescent lamps . . . 36,133
 TRAPATT oscillator . . . 40, 99
 Travelling-wave divider for UHF . . . 38, 54
 TTC (Thermal-Transient Compensated) cut for quartz crystals . . . 40, 1
 Tungsten transport by water cycle . . . 36,133

Undercutting in wet-chemical etching . . . 38,149

Vacuum technology, cryopumps in . . . 39,246
 Video disc, see LaserVision
 Video tape, making tracks visible . . . 40,129
Video telephony:
 experimental system . . . 36, 85
 digital circuits in video telephone . . . 36,233
 signal switching . . . 36,291
 Viewdata filter, integrated, with switched capacitors . . . 41,105

Visual conspicuity, measurement of . . . 36, 71
 VLP, see LaserVision
VLSI:
 negative electron resists . . . 39,346
 for Compact Disc* . . . 40,173
 X-ray lithography for . . . 41,150
 optical aspects of Silicon Repeater . . . 41,268
 Von Mises stress, determination from neck of test specimen . . . 42, 11
 Vulcanization of silicon rubbers, polymer chemistry of . . . 42,149

Water cycle, transport of tungsten by . . . 36,133
 Water supply, control of treatment plant for . . . 36,273
 Wavelength-division multiplex, optical communication system with . . . 42,245
 Wave soldering electronic components . . . 38,135
 WDM (wavelength-division multiplex), optical communication system with . . . 42,245
Wear-resistant coatings:
 deposition on tool steel by CVD . . . 40,204
 low-friction, application by PVD . . . 41,186
 Weighing and sorting machine . . . 42,173
Welding:
 laser, beam manipulation with optical fibres . . . 42,262
 of lead-through wires* . . . 42,325
 Window for ultrasoft X-rays from space . . . 40, 12

X-rays:
 hard, channel-plate image intensifier for . . . 37,124
 crystal structure research at high pressures with PW 1100 . . . 38,246
 digital image enhancement . . . 38,298
 flashing tomosynthesis . . . 38,338

computerized mammogram processing . . . 38,347
 electroradiography for medical X-ray diagnosis . . . 39, 19
 window for ultrasoft X-rays from space . . . 40, 12
 TOMOSCAN 310* . . . 40,253
 metal/ceramic tubes for non-destructive testing . . . 41, 24
 imaging with Compton-scatter radiation . . . 41, 46
 metal/ceramic diagnostic tube . . . 41,126
 image intensifier with large input format . . . 41,137
 lithography for VLSI . . . 41,150
 diffractometer, high-resolution . . . 41,183
 scandium for tube anode* . . . 41,267
 diagnostics, dual-energy . . . 42,274
 medical systems in the last half-century . . . 42,352
 diagnostic tube with spiral-groove bearings . . . 44,357

Young Scientists and Inventors:

special issue . . . { 38, 1
 -39
 preface . . . 38, 1
 inquiry . . . 38, -2
 pollution of the Shannon in Limerick . . . 38, 6
 lichens . . . 38, 8
 hovercraft sprayer . . . 38, 11
 local functional equations . . . 38, 13
 instrument for measuring cave cross-section . . . 38, 14
 lecture 'Science and Education' . . . 38, 17
 lecture 'Science and Society' . . . 38, 25
 lecture 'Science and World Problems' . . . 38, 33

Zinc-diffusion profiles in GaP and GaAsP . . . 37,121

Author index, Volumes 36-44

Figures in bold type indicate the volume number, and those in ordinary type the page number. Articles published in volumes 1-35 are given in the author indexes at the end of volumes 10, 15, 20, 25, 30, 35 and 40.

- Aagaard, E. A.**, P. M. van den Avoort and F. W. de Vrijer
An experimental video-telephone system 36, 85
- Acket, G. A.**, J. J. Daniele, W. Nijman, R. P. Tijburg and P. J. de Waard
Semiconductor lasers for optical communication 36,190
- Admiraal, D. J. H.**, B. L. Cardozo, G. Domburg and J. J. M. Neelen
Annoyance due to modulation noise and drop-outs in magnetic sound recording 37, 29
- Akkerman, H. J.**
Engineering technology (*35 years of IKO*) 39,315
- Aldefeld, B.**
Calculation and design of electro-mechanical devices 39, 78
- Alphen, M. P. van**, R. E. J. van de Grift, J. M. Pieper and R. J. van de Plassche
The PM 2517 automatic digital multimeter 38,181
- Alphen, W. M. van**, see Spanjer, T. G.
- Amato, M.**, G. Bruning, S. Mukherjee and I. T. Wacyk
Power integrated circuits 44,310
- André, J. P.**, see Frijlink, P. M.
- Annegarn, M. J. J. C.**, J. P. Arragon, G. de Haan, J. H. C. van Heuven and R. N. Jackson
HD-MAC: a step forward in the evolution of television technology 43,197
- , A. H. H. J. Nillesen and J. G. Raven
Digital signal processing in television receivers 42,183
- Arragon, J. P.**, see Annegarn, M. J. J. C.
- Asjes, R. J.**, C. S. Caspers and C. H. F. Velzel
A laser-Doppler displacement meter 43,180
- Ass, H. M. J. M. van**, P. Geittner, R. G. Gossink, D. Küppers and P. J. W. Severin
The manufacture of glass fibres for optical communication 36,182
- Asselman, G. A. A.** and A. J. van Mensfoort
A refrigerator-freezer with heat pipe 40,350
- Auphan, M.** and G. Dale
A transducer for medical echography 38,195
- Avoort, P. M. van den**, see Aagaard, E. A.
- Bacchi, H.** and A. Moreau
Real-time orthogonal transformation of colour-television pictures 38,119
- Baig, W. G.**
An edge-enhancing double-focus camera for image processing 37,180
- Bakker, P.**
Linear electron accelerators (*35 years of IKO*) 39,325
- Bartels, W. J.**
High-resolution X-ray diffractometer 41,183
- Barth, P. J.**, see Voorman, J. O.
- Bastiaens, J. J. J.** and W. C. H. Gubbels
256-kbit SRAM: an important step on the way to sub-micron IC technology 44, 33
- Baudet, P.**, M. Binet and D. Boccon-Gibod
Low-noise microwave GaAs field-effect transistor 39,269
- Beasley, J. P.** and D. G. Squire
Electron-beam pattern generator 37,334
- Beek, L. K. H. van**
Polymer chemistry in the electrical industry 42,149
- Beenakker, C. I. M.**, P. W. J. M. Boumans and P. J. Rommers
A microwave-induced plasma as an excitation source for atomic emission spectrometry 39, 65
- Behr, J.-P.**, P. Pernaut, B. Schendel and J. Schwandt
Modelling and simulation as an aid in designing a computer 39,134
- Beireus, L. C. M.** and A. A. van Gorkum
Distortion due to spherical aberration 42, 20
- Belouet, C.**
DKDP crystals for use in the TITUS tube 39, 50
- Berg, J. F. M. van de**, T. E. G. Daenen, G. Krijl and R. E. van de Leest
The electrodeposition of aluminium 39, 87
- Berkhout, P. J.**, see Eggermont, L. D. J.
- Bethe, K.** and D. Schön
Thin-film strain-gauge transducers 39, 94
- Binet, M.**, see Baudet, P.
- Blaffert, T.**
EXPERTISE: an expert system for infrared spectrum evaluation 44, 44
- Bleeker, J. A. M.**, W. H. Diemer, A. P. Huben and H. Huizenga
Camera window for ultrasoft X-rays from celestial sources 40, 12
- Bloem, H.**, J. C. de Grijs and R. L. C. de Vaan
An evacuated tubular solar collector incorporating a heat pipe 40,181
- Bloem, J.** and W. A. P. Claassen
Nucleation and growth of silicon films by chemical vapour deposition 41, 60
- , A. Bouwknecht and G. A. Wesselink
Amalgams for fluorescent lamps 38, 83
- Blood, P.**, C. T. Foxon and E. D. Fletcher
The application of semiconductor superlattices to short-wavelength lasers 44,268
- , see Kucharska, A. I.
- Blume, P.**
Computer-aided design 36,162
- Boccon-Gibod, D.**, see Baudet, P.
- Boef, A. J. den**, see Verbeek, B. H.
- Bollen, L. J. M.**, J. J. Goedbloed and E. T. J. M. Smeets
The avalanche photodiode 36,205
- Bosma, H.** and W. G. Gelling
LSI—a revolution in electronics 37,267
- Botden, P. J. M.**, see Kramer, C.
- Boudewijus, H. P. J.**, E. C. Dijkmans, P. W. Millenaar, N. A. M. Verhoeckx and C. H. J. Vos
Digital circuits in the video telephone 36,233
- Boulou, M.**, M. Furtado and G. Jacob
Light-emitting diodes based on GaN 37,237
- Boumaus, P. W. J. M.**, see Beenakker, C. I. M.
- Boutot, J. P.** and J. C. Delmotte
Two microchannel-plate photomultipliers with subnanosecond characteristics 38,240
- Bouwer, A. G.**, G. Bouwhuis, H. F. van Heek and S. Wittekoek
The Silicon Repeater 37,330
- Bouwhuis, G.**, see Bouwer, A. G.
- Bouwknegt, A.**, see Bloem, J.
- Braat, J. J. M.**
Aspherics, II. Aspheric surfaces: design and optical advantages 41,289
- , see Hartmann, M.
- Brandsma, J. R.**
PHILAN, a local-area network based on a fibre-optic ring 43, 10
- Breed, D. J.**, F. H. de Leeuw, W. T. Stacy and A. B. Voermans
Garnet films for fast magnetic bubbles 38,211
- Brehm, R.**, K. van Dun, J. Haisma and J. C. G. Teunissen
Transparent single-point turning of glass 39, 92
- Brice, J. C.**, M. J. Hight, O. F. Hill and P. A. C. Whiffin
Pulling large bismuth-silicon-oxide crystals 37,250
- and W. Koelewijn
Selecting quartz for resonators 43,214
- and W. S. Metcalf
Quartz-crystal resonators using an unconventional cut 40, 1

- Brinkman, G. A.**
Radioisotopes (*35 years of IKO*) 39,294
- Broek, C. A. M. van den and A. L. Stuijts**
Ferroxudure 37,157
- Broek, J. J. van den and A. G. Dirks**
Metastable phases and thermodynamic equilibrium 43,304
- Broese van Groenou, A. and J. D. B. Veldkamp**
Grinding brittle materials 38,105
- Brouha, M. and A. G. Rijnbeek**
Apparatus for solid-state research at very high pressures 36,245
- Brouwer, J. F.**
Half a century of 'electronification' in telephony systems 42,361
- Bruffaerts, A., E. Henin and A. Pirotte**
A sound basis for the generation of explanations in Expert Systems 44,287
- Bruning, G., see Amato, M.**
- Bulle-Lieuwma, C. W. T., see Zalm, P. C.**
- Bulthuis, K.**
A farewell message 44,325
- Bunge, E.**
Speaker recognition by computer 37,207
- Burnett, D. J.**
INDA, a software tool for the production engineer 40,237
- Buschow, K. H. J.**
Research on amorphous alloys 42, 48
- Carasso, M. G., J. B. H. Peek and J. P. Sinjou**
The Compact Disc Digital Audio System 40,151
- Cardozo, B. L., see Admiraal, D. J. H.**
- Carl, K., J. A. M. Dikhoff and W. Eckenbach**
The pigmentation of phosphors for colour television 40, 48
- Casimir, H. B. G.**
Gilles Holst, pioneer of industrial research in the Netherlands 40,121
Marginal notes for an anniversary 42,295
- Caspers, C. S., see Asjes, R. J.**
- Chalmeton, V.**
A channel-plate image intensifier for hard X-rays 37,124
- Christiaens, M., see Jager, F. de**
- Claassen, W. A. P., see Bloem, J.**
- Clarke, J. A., see Vriens, L.**
- Coehoorn, R., see Schuurmans, M. F. H.**
- Coenders, J. W.**
Switching telephone and video-telephone signals 36,291
- Conner, G. and R. H. Lane**
HS3: an advanced bipolar-IC technology 44,296
- Coppelmans, P. M. C., see Roermund, A. H. M. van**
- Coumans, J. J. H.**
Dual-energy X-ray diagnostics 42,274
- Cremers, R. M. M., see Weijer, P. van de**
- Crucq, J.**
Theory and practice of acoustic noise control in electrical appliances 44,123
- Daenen, T. E. G., see Berg, J. F. M. van de**
- Dale, G., see Auphan, M.**
- Daniele, J. J., see Acket, G. A.**
- Daniels, J., K. H. Härdtl and R. Wernicke**
The PTC effect of barium titanate 38, 73
- Dantzig, R. van**
BOL (*35 years of IKO*) 39,302
- Davies, R., B. H. Newton and J. G. Summers**
The TRAPATT oscillator 40, 99
- Davis, G. L.**
Transport of tungsten by the water cycle 36,133
- De Man, H., see Meerbergen, J. L. van**
- Dekker, W.**
Bright spots and bottlenecks in Europe's future industrial development 43, 77
- Dekkers, N. H. and H. de Lang**
A detection method for producing phase and amplitude images simultaneously in a scanning transmission electron microscope 37, 1
- Delmotte, J. C., see Boutot, J. P.**
- Denner, W. and Heinz Schulz**
Apparatus based on Philips PW 1100 diffractometer for crystal-structure research at high pressures 38,246
- Diemer, W. H., see Bleeker, J. A. M.**
- Dijken, R. H.**
Electric motors in small domestic appliances 42,336
- Dijkmans, E. C., see Boudewijns, H. P. J.**
- Dijksman, J. F.**
Analysis of the injection-moulding process 44,212
- Dikhoff, J. A. M., see Carl, K.**
- Dimigen, H. and H. Hübsch**
Applying low-friction wear-resistant thin solid films by physical vapour deposition 41,186
—, see Kobs, K.
- Dirks, A. G., see Broek, J. J. van den**
- Dixon, Bernard**
Science and Education 38, 17
- Dolizy, P.**
Growth of alkali-antimonide films for photocathodes 40, 19
- Dollekamp, H., L. J. M. Esser and H. de Jong**
P²CCD in 60MHz oscilloscope with digital image storage 40, 55
- Dolphin, R. J., L. P. J. Hoogeveen and F. W. Willmott**
An experimental system for the automatic determination of organochlorine residues in milk 36,284
- Domburg, G., see Admiraal, D. J. H.**
- Doorn, R. A. van and N. A. M. Verhoeckx**
An I²L digital modulation stage for data transmission 37,291
- Döring, M.**
Ink-jet printing 40,192
- Dössel, O., M. H. Kuhn and H. Weiss**
Magnetic fields in medical diagnostics: MR and SQUID 44,259
- Dötsch, H.**
The microwave generation and manipulation of magnetic domains 37, 38
- Druyvesteyn, W. F., F. A. Kuijpers, A. G. H. Verhulst and C. H. M. Witmer**
Single-mask bubble memory with rotating field-control 36,149
- Dun, K. van, see Brehm, R.**
- Eckenbach, W., see Carl, K.**
- Eggermont, L. D. J. and P. J. Berkhout**
Digital audio circuits: computer simulations and listening tests 41, 99
—, M. H. H. Höfelt and R. H. W. Salters
A delta-modulation to PCM converter 37,313
- Eijk, B. van der and W. Kühl**
An X-ray image intensifier with large input format 41,137
- Elst, J. H. R. M. and D. K. Wielenga**
The finite-element method and the ASKA program, applied in stress calculations for television picture tubes 37, 56
- Enden, A. W. M. van den and N. A. M. Verhoeckx**
Digital signal processing: theoretical background 42,110
- Engel, F. L.**
The measurement of visual conspicuity 36, 71
- Engelen, G. A. J. van, J. L. M. Hagen and W. A. L. Heijnemans**
An equipment for measuring the magnetic fields of television deflection coils 39,277
- Engelsen, D. den, J. H. Th. Hengst and E. P. Honig**
An automated Langmuir trough for building monomolecular layers 36, 44
- Engelsma, G.**
Phenol synthesis and photomorphogenesis 38, 89
- Eppenga, R. and M. F. H. Schuurmans**
Theory of the GaAs/AlGaAs quantum well 44,137
—, see Schuurmans, M. F. H.
- Erman, M., see Frijlink, P. M.**
- Esser, L. J. M., see Dollekamp, H.**
- Finck, J. C. J., H. J. M. van der Laak and J. T. Schrama**
A semiconductor laser for information read-out 39, 37
- Fischer, W. E.**
A data-base management system for CAD and CAM 40,245
- Fletcher, E. D., see Blood, P.**
—, see Kucharska, A. I.
- Foederer, A. F., J. L. M. Hagen and A. G. van Nie**
An instrument for measuring the curvature of reflecting surfaces 40,338
- Foxon, C. T., see Blood, P.**
—, see Joyce, B. A.
- Frank, G., see Köstlin, H.**
- Franken, A. J. J., G. D. Khoe, J. Renkens and C. J. G. Verwer**
Experimental semi-automatic machine for hot splicing glass fibres for optical communication 38,158
—, see Wierenga, P. E.
- Frankfort, F. J. M.**
Vibration patterns and radiation behaviour of loud-speaker cones 36, 1
- Fransen, J. J. B. and J. H. N. van Vucht**
An easily controlled alkali-vapour dispenser 36, 16

- French, R. C. and P. J. Mabey
Error control in mobile-radio data communication . . . 39,172
- Frens, G., H. F. Huisman, J. K. Vondeling and K. M. van der Waarde
Suspension technology 36,264
- Frijlink, P. M., J. P. André and M. Erman
Metal-organic vapour-phase epitaxy of multilayer structures with III-V semiconductors 43,118
- Furtado, M., see Boulou, M.
- Galenkamp, H. and H. van Wijngaarden
Determining the Von Mises stress from the neck of a tensile-test specimen 42, 11
- Geittner, P. and H. Lydtin
Manufacturing optical fibres by the PCVD process . . . 44,241
—, see Ass, H. M. J. M. van
- Gelling, W. G. and F. Valster
The new centre for submicron IC technology 42,266
—, see Bosma, H.
- Gerwen, P. J. van, W. A. M. Sniijders and N. A. M. Verhoeckx
An integrated echo canceller for baseband data transmission 39,102
- Gestel, W. J. van, F. W. Gorter and K. E. Kuijk
Read-out of a magnetic tape by the magnetoresistance effect 37, 42
- Giellis, G. C. M., J. B. H. Peek and J. M. Schmidt
Station and programme identification in FM sound broadcasting 39,216
- Gijsbers, T. G.
COLATH, a numerically controlled lathe for very high precision 39,229
—, see Haisma, J.
- Goedbloed, J. J., see Bollen, L. J. M.
- Goedhart, D., R. J. van de Plassche and E. F. Stikvoort
Digital-to-analog conversion in playing a Compact Disc 40,174
- Gorkom, G. G. P. van and A. M. E. Hoeberechts
Silicon cold cathodes 43, 49
- Gorkum, A. A. van, see Beirens, L. C. M.
—, see Spanjer T. G.
- Gorter, F. W., see Gestel, W. J. van
- Gossink, R. G., see Ass, H. M. J. M. van
- Graeger, V., R. Kobs and M. Liehr
A ceramic differential-pressure transducer 43, 86
- Gravesteijn, D. J. and J. van der Veen
Organic-dye films for optical recording 41,325
—, C. J. van der Poel, P. M. L. O. Scholte and C. M. J. van Uijen
Phase-change optical recording 44,250
- Grift, R. E. J. van de, see Alphen, M. P. van
- Grijs, J. C. de, see Bloem, H.
- Groh, G.
The challenge of picture processing 38,291
- Gross, U., F. J. M. Mescher and J. C. Tiemeijer
The microprocessor-controlled CM12/STEM scanning-transmission electron microscope 43,273
- Gubbels, W. C. H., see Bastiaens, J. J. J.
- Guildford, L. H.
The Dot Scan CCTV, a flexible system for real-time image-processing experiment 38,310
— and B. D. Young
A mobile system for image bulk storage 43,260
- Haan, G. de, see Annegarn, M. J. J. C.
- Hagemann, H.-J., D. Hennings and R. Wernicke
Ceramic multilayer capacitors 41, 89
- Hagen, J. L. M., see Engelen, G. A. J. van
—, see Foederer, A. F.
- Haisma, J. and T. G. Gijsbers
Aspherics, I. Optomechanics, an ultra-high-precision machining technique 41,285
—, W. Mesman, J. M. Oomen and J. C. Wijn
Aspherics, III. Fabrication, testing and application of highly accurate aspheric optical elements 41,296
—, see Brehm, R.
- Hanenberg, J. G. van den and J. Vredenburg
An experimental assembly robot 40, 33
- Hansen, P., B. Hill and W. Tolksdorf
Optical switching with bismuth-substituted iron garnets 41, 33
- Harding, G., H. Strecker and R. Tischler
X-ray imaging with Compton-scatter radiation 41, 46
- Härdtl, K. H. see Daniels, J.
- Harrop, P., P. Lesarte and T. H. A. M. Vlek
Low-noise 12GHz front-end designs for direct satellite television reception 39,257
- Hart, J. 't, S. G. Nootboom, L. L. M. Vogten and L. F. Willems
Manipulation of speech sounds 40,134
- Hartl, W., D. Peter and K. Reiber
Metal/ceramic X-ray tubes for non-destructive testing . 41, 24
A metal/ceramic diagnostic X-ray tube 41,126
- Hartmann, M., B. A. J. Jacobs and J. J. M. Braat
Erasable magneto-optical recording 42, 37
- Haverkorn van Rijsewijk, H. C., P. E. J. Legierse and G. E. Thomas
Manufacture of LaserVision video discs by a photopolymerization process 40,287
- Hazan, J. P. and L. Jacomme
Characterizing optical fibres; a test bench for pulse dispersion 36,211
- Heek, H. F. van, see Bouwer, A. G.
- Heemskerk, J. P. J. and K. A. Schouhamer Immink
Compact Disc: system aspects and modulation 40,157
- Heijden, J. van der
DIVAC — an experimental optical-fibre communications network 41,253
- Heijman, M. G. J., J. H. W. Kuntzel and G. H. J. Somers
Multi-track magnetic heads in thin-film technology . . 44,169
- Heijnenmans, W. A. L., J. A. M. Nieuwendijk and N. G. Vink
The deflection coils of the 30AX colour-picture system 39,154
—, see Engelen, G. A. J. van
- Heitmann, H., B. Hill, J.-P. Krumme and K. Witter
MOPS, a magneto-optic storage wafer of the discrete-bit type 37,197
- Hemert, J. P. van
Automatic segmentation of speech into diphones . . . 43,233
- Hengst, J. H. Th., see Engelsen, D. den
- Henin, E., see Bruffaerts, A.
- Hennings, D., see Hagemann, H.-J.
—, see Lütteke, G.
- Heusen, S. van and L. G. J. Mans
Measurement of ozone in an aircraft 38,131
- Heuvel, F. C. van den
Striations in a gas discharge 44, 89
- Heuven, J. H. C. van, see Annegarn, M. J. J. C.
- Heyns, H., H. L. Peek and J. G. van Santen
Image sensor with resistive electrodes 37,303
- Hight, M. J., see Brice, J. C.
- Hill, B., see Hansen, P.
—, see Heitmann, H.
- Hill, O. F., see Brice, J. C.
- Hily, C., J. J. Hunzinger, M. Jatteau and J. Ott
Real-time macro- and microthermography 37,241
- Hoeberechts, A. M. E., see Gorkom, G. G. P. van
- Hoeve, H., J. Timmermans and L. B. Vries
Error correction and concealment in the Compact Disc system 40,166
- Höfelt, M. H. H., zie Eggermont, L. D. J.
- Hofker, W. K.
Geiger-Müller counters (35 years of IKO) 39,296
Semiconductor detectors (35 years of IKO) 39,298
Electronics for nuclear physics (35 years of IKO) . . . 39,312
Ion-beam technology (35 years of IKO) 39,320
— and J. Politiek
Ion implantation in semiconductors 39, 1
- Holster, P. L., J. A. H. M. Jacobs and B. Sastra
Measuring the radial error of precision air bearings . . 41,334
- Honds, L. and K. H. Meyer
A linear d.c. motor with permanent magnets 40,329
- Honig, E. P., see Engelsen, D. den
- Hoof, G. W. 't, see Leys, M. R.
- Hoogeveen, L. P. J., see Dolphin, R. J.
- Hoppe, W. J. J. van, G. D. Khoe, G. Kuyt and H. F. G. Smulders
Very smooth fracture faces for optical glass fibres . . . 37, 89
- Houten, S. van
Applied research — the source of innovation in consumer electronics 44,180
- Howden, H.
Production of optical correction plates for projection television 39, 15

- Hoyer, A. and M. Schlindwein
Digital image enhancement 38,298
— and W. Spiesberger
Computerized mammogram processing 38,347
- Huben, A. P., see Bleeker, J. A. M.
- Hübsch, H., see Dimigen, H.
—, see Kobs, K.
- Huisman, H. F. and C. J. F. M. Rasenberg
Porosimeter measurements on magnetic tape 41,260
—, see Frens, G.
- Huizing, A., A. H. T. Sanders and J. F. K. Thijssen
Weighing and sorting machine 42,173
- Huizinga, H., see Bleeker, J. A. M.
- Hunzinger, J. J., see Hily, C.
- Immink, K. A. Schouhamer, see Heemskerk, J. P. J.
- Jack, A. G. and Q. H. F. Vrehan
Progress in fluorescent lamps 42,342
- Jackson, R. N., see Annegarn, M. J. J. C.
- Jacob, G., see Boulou, M.
- Jacobs, B. A. J., see Hartmann, M.
—, see Vriens, L.
- Jacobs, C. A. J. and J. A. J. M. van Vliet
A new generation of high-pressure sodium lamps 39,211
- Jacobs, J. A. H. M., see Holster, P. L.
- Jacomme, L., see Hazan, J. P.
- Jager, F. de and M. Christiaens
A fast automatic equalizer for data links 37, 10
- Jager, K., see Wit, H. J. de
- Jagt, J. C. and P. W. Whipps
Negative electron resists for VLSI 39,346
- Janssen, R. K., see Krol, D. M.
- Jatteau, M., see Hily, C.
- Jeu, W. H. de and J. van der Veen
Liquid crystals for numerical displays 37,131
- Jong, H. de, see Dollekamp, H.
- Jonge, A. K. de, see Stolff, F.
- Joyce, B. A. and C. T. Foxon
Molecular beam epitaxy of multilayer structures with
GaAs and $Al_xGa_{1-x}As$ 43,143
- Kalter, H. and E. P. G. T. van de Ven
Plasma etching in IC technology 38,200
- Kam, J. J. van der
A digital 'decimating' filter for analog-to-digital conversion of hi-fi audio signals 42,230
- Kamerbeek, E. M. H.
Magnetic bearings 41,348
- Kanters, J. T., see Peters, J. H.
- Kasperkovitz, W. D.
Frequency-dividers for ultra-high frequencies 38, 54
- Kasperkovitz, W. G.
FM receivers for mono and stereo on a single chip 41,169
- Kats, A.
Glass — outline of a development 42,316
- Kelly, J. J. and G. J. Koel
Galvanic effects in the wet-chemical etching of metal films 38,149
—, P. H. L. Notten, J. E. A. M. van den Meerakker and R. P. Tijburg
Wet-chemical etching of III-V semiconductors 44, 61
- Kelly, P. J., see Schuurmans, M. F. H.
- Kessels, J. L. W. and A. J. Martin
Parallel programs 40,254
- Khoe, G. D. and L. J. Meuleman
Light modulation and injection in optical-fibre transmission systems with semiconductor lasers 36,201
—, see Franken, A. J. J.
—, see Hoppe, W. J. J. van
- Kilian, R. and M. Liehr
Experimental etching equipment 38, 51
- Klein Wassink, R. J.
The thermal behaviour of electronic components during soldering 38,135
— and H. J. Vledder
The attachment of leadless components to printed boards 40,342
- Klinck, M.
Control of the surface-water purification plant for the Amsterdam Water-Supply Authority 36,273
- Kloosterboer, J. G., G. J. M. Lippits and H. C. Meinders
Photopolymerizable lacquers for LaserVision video discs 40,298
- Klotz, E., R. Linde, U. Tiemens and H. Weiss
Flashing tomosynthesis 38,338
- Kluitmans, J. T. M., see Tjassens, H.
- Knippenberg, W. F. and B. Lersmacher
Carbon foam 36, 93
—, B. Lersmacher and H. Lydtin
Products of pyrolytic graphite 37,189
- Knowles, J. E.
Measuring the magnetic properties of a microscopic particle 39, 48
- Kobs, K., H. Dimigen, H. Hübsch and H. J. Tolle
Improved adhesion of solid lubricating films with ion-beam mixing 44, 24
- Kobs, R., see Graeger, V.
- Koel, G. J., see Kelly, J. J.
- Koelwijjn, W., see Brice, J. C.
- Koning, S. H. de
The MCR system — multiple-channel amplification of reverberation 41, 12
- Köstlin, H. and G. Frank
Thin-film reflection filters 41,225
—, see Schaper, H.
- Kramer, C. and P. J. M. Botden
Medical systems in the last half century 42,352
- Kramer, P.
75 years of research: from lamps to integrated circuits 44,239
- Krijl, G., see Berg, J. F. M. van de
- Krol, D. M. and R. K. Janssen
Research on television glass 43,253
- Krol, Th.
The '(4,2) concept' fault-tolerant computer 41, 1
- Krumme, J.-P., see Heitmann, H.
- Kruseman Aretz, F. E. J.
Abstraction 40,225
- Kucharska, A. I., P. Blood and E. D. Fletcher
Bistability in quantum-well lasers 44, 76
- Kühl, W., see Eijk, B. van der
- Kuhn, M. H. and W. Menhardt
Interactive MR image synthesis 43, 95
—, see Dössel, O.
- Kuijk, K. E., see Gestel, W. J. van
- Kuijpers, F. A., see Druyvesteyn, W. F.
- Kuntzel, J. H. W., see Heijman, M. G. J.
- Küppers, D., see Ass, H. M. J. M. van
- Kurz, H.
Lithium niobate as a material for holographic information storage 37,109
- Kuyt, G., see Hoppe, W. J. J. van
- Laak, H. J. M. van der, see Finck, J. C. J.
- Lane, R. H., see Conner, G.
- Lang, H. de, see Dekkers, N. H.
- Leest, R. E. van de, see Berg, J. F. M. van de
- Leeuw, F. H. de, see Breed, D. J.
- Legierse, P. E. J., see Haverkorn van Rijsewijk, H. C.
- Lens, G. A., see Zaengel, T.
- Lenstra, D., see Verbeek, B. H.
- Lersmacher, B., see Knippenberg, W. F.
- Lesarte, P., see Harrop, P.
- Leys, M. R., M. P. A. Viegiers and G. W. 't Hooff
Metal-organic vapour-phase epitaxy with a novel reactor and characterization of multilayer structures 43,133
- Liehr, M., see Graeger, V.
—, see Kilian, R.
- Ligthart, H. J. and J. Politiek
An open 800-kV ion-implantation machine 43,169
- Linde, R., see Klotz, E.
- Lippits, G. J. M., see Kloosterboer, J. G.
- Locher, P. R.
Proton NMR tomography 41, 73
- Lorentz, G.
Technological aspects of advanced telecommunications 44, 16
- Luijckx, G.
The cyclotron (35 years of IKO) 39,290
- Lüthje, H.
X-ray lithography for VLSI 41,150
- Lütteke, G. and D. Hennings
Dielectric resonators for microwave integrated oscillators 43, 35
- Luyt, B. A. G. van and L. E. Zegers
The Compact Disc Interactive System 44,326
- Lydtin, H., see Geittner, P.
—, see Knippenberg, W. F.
- Mabey, P. J., see French, R. C.
- Man, H. De, see Meerbergen, J. L. van
- Mans, L. G. J., see Heusden, S. van

- Marée, P. M. J., see Zalm, P. C.
- Martienssen, W.
Science and Society 38, 25
- Martin, A. J.
Distributed computations on arrays of processors 40,270
—, see Kessels, J. L. W.
- Mauczok, R. and R. Wernicke
Ceramic boundary-layer capacitors 41,338
- Meerakker, J. E. A. M. van den, see Kelly, J. J.
- Meerbergen, J. L. van
Developments in integrated digital signal processors, and the PCB 5010 44, 1
— and H. De Man
A true silicon compiler for the design of complex ICs for digital signal processing 44,218
- Meerman, W. C. P. M., T. L. van Rooy and M. C. M. Voss
Continuous skull-melting of glass 42, 93
- Meijer, E. W.
Polypeoxides; formation and properties of their network structure 44,110
- Meijer, F.
Looking at interfaces 43,220
- Meinders, H. C., see Kloosterboer, J. G.
- Melis, J. H. A.
O-BUS: a system for flexible public transport by means of on-call buses 40,231
- Memming, R.
Energy production by photoelectrochemical processes 38,160
- Menhardt, W., see Kuhn, M. H.
- Mensvoort, A. J. van, see Asselman, G. A. A.
- Mescher, F. J. M., see Gross, U.
- Mesman, W., see Haisma, J.
- Metcalf, W. S., see Brice, J. C.
- Meuleman, L. J., see Khoe, G. D.
- Meyer, K. H., see Hondts, L.
- Meyer, W. and W. Schilz
Microwave measurement of moisture content in process materials 40,112
- Michel, C.
Observations of domains in ferroelectrics and ferro-magnetics with a scanning electron microscope 36, 18
- Miedema, A. R.
The heat of formation of alloys 36,217
The atom as a metallurgical building block 38,257
Layered semiconductor structures, Editorial 43,109
- Millenaar, P. W., see Boudewijns, H. P. J.
- Moreau, A., see Bacchi, H.
- Mouthaan, K.
Optical communication systems with glass-fibre cables 36,178
- Muijderman, E. A., G. Remmers and L. P. M. Tielemans
Grease-lubricated spiral-groove bearings 39,184
—, C. D. Roelandse, A. Vetter and P. Schreiber
A diagnostic X-ray tube with spiral-groove bearings 44,357
- Mukherjee, S., see Amato, M.
- Mulder, B. J. and R. van der Schee
A piezoelectric microbalance 41,304
- Neelen, J. J. M., see Admiraal, D. J. H.
- Newton, B. H., see Davis, R.
- Nicia, A. J. A.
An optical communication system with wavelength-division multiplexing and minimized insertion losses, I. System and coupling efficiency 42,245
— and C. J. T. Potters
Components for glass-fibre circuits 40, 46
—, C. J. T. Potters and A. H. L. Tholen
An optical communication system with wavelength-division multiplexing and minimized insertion losses, II. Multiplexing and demultiplexing 43,344
- Nie, A. G. van
A method of measuring mechanical stresses in passivation layers 39,130
—, see Foederer, A. F.
- Niessen, C.
Computer-aided design of LSI circuits 37,278
- Nieuwendijk, J. A. M., see Heijnemans, W. A. L.
- Nieuwland, J. M. van, A. Petterson and C. Weber
The design and construction of a non-rectangular reverberation chamber 37,176
- Nijman, W., see Acket, G. A.
- Nillesen, A. H. H. J., see Annegarn, M. J. J. C.
- Nonhof, C. J. and G. J. A. M. Notenboom
Beam manipulation with optical fibres in laser welding 42,262
- Nooteboom, S. G., see Hart, J. 't
- Notenboom, G. J. A. M., see Nonhof, C. J.
- Notten, P. H. L., see Kelly, J. J.
- Obertop, D., see Willich, P.
- Oomen, J. M., see Haisma, J.
- Oostrum, K. J. van
CAD in light optics and electron optics 42, 69
- Opdorp, C. van, see Werkhoven, C.
- Osenbruggen, C. van and C. de Regt
Electrochemical micromachining 42, 22
- Ott, J., see Hily, C.
- Pannenberg, A. E.
Science and World Problems 38, 33
- Parker, D. W., R. G. Pratt, F. W. Smith and R. Stevens
Acoustic surface-wave bandpass filters 36, 29
- Peek, H. L., see Heyns, H.
- Peek, J. B. H.
Digital signal processing — growth of a technology 42,103
—, see Carasso, M. G.
—, see Gielis, G. C. M.
- Peelen, J. G. J.
Light transmission of sintered alumina 36, 47
—, B. V. Rejda and J. P. W. Vermeiden
Sintered hydroxylapatite as a bioceramic 37,234
- Pernards, P., see Behr, J.-P.
- Persoon, E. H. J.
A system that can learn to recognize two-dimensional shapes 38,356
— and C. J. B. Vandembulcke
Digital audio: examples of the application of the ASP integrated signal processor 42,201
- Peschmann, K. R.
Medical electroradiography — its potential and limitations 39, 19
- Peter, D., see Hartl, W.
- Peters, J. H. and J. T. Kanters
CAROT: a digital method of increasing the robustness of an analog colour television signal 42,217
- Petersen, A., P. Schnabel, H. Schweppe and R. Wernicke
A small analog memory based on ferroelectric hysteresis 37, 51
- Petterson, A., see Nieuwland, J. M. van
- PHLIQA Project Group
PHLIQA 1, a question-answering system for data-base consultation in natural English,
I. Organisation and performance 38,229
II. The artificial languages and translation operations 38,269
- Pieper, J. M., see Alphen, M. P. van
- Pirotte, A., see Bruffaerts, A.
- Plassche, R. J. van de, see Alphen, M. P. van
—, see Goedhart, D.
- Poel, C. J. van der, see Gravesteyn, D. J.
- Politiek, J., see Hofker, W. K.
—, see Ligthart, H. J.
- Ponjeé, J. J. and P. N. T. van Velzen
Chemical modification in surfaces 44, 81
- Potters, C. J. T., see Nicia, A. J. A.
- Prast, G.
Quantitative measurements by the Schlieren method 43,184
- Pratt, R. G., see Parker, D. W.
- Rasenberg, C. J. F. M., see Huisman, H. F.
- Raue, R., A. T. Vink and T. Welker
Phosphor screens in cathode-ray tubes for projection television 44,335
- Raven, J. G., see Annegarn, M. J. J. C.
- Regt, C. de, see Osenbruggen, C. van
- Reiber, K., see Hartl, W.
- Rejda, B. V., see Peelen, J. G. J.
- Remmers, G., see Muijderman, E. A.
- Renkens, J., see Franken, A. J. J.
- Rijckaert, A. M. A.
Making the tracks on video tape visible with a magnetic fluid 40,129
- Rijnbeek, A. G., see Brouha, M.
- Rocchi, M.
Research on monolithic GaAs MESFET circuits at LEP 44,302
- Roelandse, C. D., see Muijderman, E. A.
- Roermund, A. H. M. van and P. M. C. Coppelmans
An integrated switched-capacitor filter for viewdata 41,105
- Rommers, P. J., see Beenakker, C. I. M.

- Roosmalen, J. H. T. van
A new concept for television camera tubes 39,201
- Rooy, T. L. van, see Meerman, W. C. P. M.
- Rothgordt, U.
Electrostatic printing 36, 57
- Salters, R. H. W., see Eggermont, L. D. J.
- Sanders, A. H. T., see Huizing, A.
- Santen, J. G. van, see Heyns, H.
- Saraga, P. and J. A. Weaver
An experiment in flexible automation 38,329
- Sastra, B., see Holster, P. L.
- Scha, R. J. H.
Software 40,219
- Schaper, H., H. Köstlin and E. Schnedler
Electrochemiluminescence in electrolyte-free solutions 40, 69
- Schee, R. van der, see Mulder, B. J.
- Scheer, J. J. and J. Visser
Application of cryopumps in industrial vacuum technology 39,246
- Schendel, B., see Behr, J.-P.
- Schiefer, G.
Broadband circulators for VHF and UHF 36,255
- Schilz, W., see Meyer, W.
- Schindwein, M., see Hoyer, A.
- Schmidt, J. M., see Gielis, G. C. M.
- Schmidt, U. J.
Electro-optic deflection of a laser beam 36,117
- Schnabel, P., see Petersen, A.
- Schnedler, E., see Schaper, H.
- Schnell, A.
Electromechanical transducers with no hysteresis 40,358
- Scholl, G. J.
A universal instrument for digital picture processing 38,326
- Scholte, P. M. L. O., see Gravesteyn, D. J.
- Schön, D., see Bethe, K.
- Schouhamer Immink, K. A., see Heemskerck, J. P. J.
- Schrama, J. T., see Finck, J. C. J.
- Schreiber, P., see Muijderland, E. A.
- Schulz, Heinz, see Denner, W.
- Schuurmans, M. F. H., R. Coehoorn, R. Eppenga en P. J. Kelly
Predicting the properties of materials: dream or reality? 44,276
—, see Eppenga, R.
- Schwandt, J., see Behr, J.-P.
- Schwappe, H., see Petersen, A.
- Scott, J. P.
Electron-image projector 37,347
- Severijns, A. P. and P. J. W. Severin
Device for stripping protective coatings from glass fibre 41,124
—, see Severin, P. J. W.
- Severin, P. J. W. and A. P. Severijns
Applications of light guides in process control 43, 58
—, see Ass, H. M. J. M. van
—, see Severijns, A. P.
- Sinjou, J. P., see Carasso, M. G.
- Sintzoff, M.
Transformation methods for improving parallel programs 40,278
- Skoyles, D. R.
A fast actuator for an anti-lock braking system 36, 74
- Sluifjerman, L. A. AE.
Chromatofocusing, a new protein-separation method 39,125
- Sluyter, R. J.
Digitization of speech 41,201
- Smeets, E. T. J. M., see Bollen, L. J. M.
- Smets, B. M. J.
On the mechanism of the corrosion of glass by water 42, 59
- Smith, F. W., see Parker, D. W.
- Smulders, H. F. G., see Hoppe, W. J. J. van
- Snijder, P. J., see Voorman, J. O.
- Snijders, W. A. M., see Gerwen, P. J. van
- Somers, G. H. J., see Heijman, M. G. J.
- Sommerdijk, J. L. and A. L. N. Stevels
The behaviour of phosphors with aluminate host lattices 37,221
- Spanjer, T. G., A. A. van Gorkum and W. M. van Alphen
Electron guns for projection television 44,348
- Spiesberger, W., see Hoyer, A.
- Spruit, J. H. M., see Vriens, L.
- Squire, D. G., see Beasley, J. P.
- Staas, F. A.
Continuous cooling in the millikelvin range 36,104
- Stacy, W. T., see Breed, D. J.
- Stevels, A. L. N., see Sommerdijk, J. L.
- Stevens, R., see Parker, D. W.
- Stikvoort, E. F., see Goedhart, D.
- Stolfi, F. and A. K. de Jonge
Stirling cryogenerators with linear drive 42, 1
- Straten, P. J. M. van der and G. Verspui
Chemical vapour deposition of wear-resistant coatings on tool steel 40,204
- Strecker, H., see Harding, G.
- Streng, J. H.
Sound radiation from a vibrating membrane 44,190
- Stuijts, A. L., see Broek, C. A. M. van den
- Summers, J. G., see Davies, R.
- Swanenburg, T. J. B.
Self-organizing systems 38,364
- Teer, K.
Looking back at distant vision: television technology from 1936 to 1986 42,297
- Teunissen, J. C. G., see Brehm, R.
- Theuwissen, A. J. P. and C. H. L. Weijters
The accordion imager, a new solid-state image sensor 43, 1
- Thijssen, J. F. K., see Huizing, A.
- Thissen, F. L. A. M.
An equipment for automatic optical inspection of connecting-lead patterns for integrated circuits 37, 77
- Tholen, A. H. L., see Nicia, A. J. A.
- Thomas, G. E.
Future trends in optical recording 44, 51
—, see Haverkorn van Rijsewijk, H. C.
- Thoone, M. L. G.
CARIN, a car information and navigation system 43,317
- Tielemans, L. P. M., see Muijderland, E. A.
- Tiemeijer, J. C., see Gross, U.
- Tiemens, U., see Klotz, E.
- Tietjens, E. W.
The history of the Philishave 42,312
- Tijburg, R. P., see Acket, G. A.
—, see Kelly, J. J.
—, see Verplanke, J. C.
- Timmermans, J., see Hoeve, H.
- Tischler, R., see Harding, G.
- Tjassens, H. and J. T. M. Kluitmans
A laser module for 4-Gbit/s optical communications 44,162
- Tolksdorf, W., see Hansen, P.
- Tolle, H. J., see Kobs, K.
- Uijen, C. M. J. van, see Gravesteyn, D. J.
- Vaan, R. L. C. de, see Bloem, H.
- Valster, F., see Gelling, W. G.
- Vandenbulcke, C. J. B., see Persoon, E. H. J.
- Veen, J. van der, see Gravesteyn, D. J.
—, see Jeu, W. H. de
- Veldhuis, J.
Computer-aided research on multiwire telephone cables 40, 85
- Veldkamp, J. D. B., see Broese van Groenou, A.
- Velzel, C. H. F., see Asjes, R. J.
- Velzen, P. N. T. van, see Ponjeé, J. J.
- Ven, E. P. G. T. van de, zie Kalter, H.
- Verbakel, J. M. M. and J. H. N. van Vucht
From powder diagram to structure model on the computer 41,239
- Verbeek, B. H., D. Lenstra and A. J. den Boef
Noise due to optical feedback in semiconductor lasers 43,292
- Verbunt, J. P. M.
Laboratory-scale manufacture of magnetic heads 44,151
- Verhoecx, N. A. M., see Boudewijns, H. P. J.
—, see Doorn, R. A. van
—, see Enden, A. W. M. van den
—, see Gerwen, P. J. van
- Verhulst, A. G. H., see Druyvesteyn, W. F.
- Vermeiden, J. P. W., see Peelen, J. G. J.
- Verplanke, J. C. and R. P. Tijburg
Determination of zinc-diffusion profiles in gallium phosphide and gallium arsenophosphide with the aid of radioactive isotopes 37,121
- Verspui, G., see Straten, P. J. M. van der
- Verweij, H.
The fining of glass 40,310
- Verwer, C. J. G., see Franken, A. J. J.
- Vessem, J. C. van
From transistor to IC: a long road? 42,326
- Vetter, A., see Muijderland, E. A.
- Viegers, M. P. A., see Leys, M. R.

- Vink, A. T., see Raue, R.
 —, see Werkhoven, C.
- Vink, N. G., see Heijnemans, W. A. L.
- Visser, J., see Scheer, J. J.
- Vitt, B.
 Black-cobalt coating for solar collectors 43,244
- Vledder, H. J., see Klein Wassink, R. J.
- Vlek, T. H. A. M., see Harrop, P.
- Vliet, J. A. J. M. van, see Jacobs, C. A. J.
- Voermans, A. B., see Breed, D. J.
- Vogten, L. L. M., see Hart, J. 't
- Volger, J.
 Cryogenics: a critical review 37, 91
- Vondeling, J. K., see Frens, G.
- Voorman, J. O., P. J. Snijder, J. S. Vromans and P. J. Barth
 An automatic equalizer for echo reduction in Teletext on a single chip 40,319
- Vos, C. H. J., see Boudewijns, H. P. J.
- Vos, J. A. de
 Megadoc, a modular system for electronic document handling 39,329
- Voss, M. C. M., see Meerman, W. C. P. M.
- Vredenburg, J., see Hanenberg, J. G. van den
- Vreken, Q. H. F., see Jack, A. G.
- Vriens, L. and B. A. J. Jacobs
 Digital optical recording with tellurium alloys 41,313
 —, J. A. Clarke and J. H. M. Spruit
 Interference filters in projection television tubes 44,201
- Vries, L. B., see Hoeve, H.
- Vrijer, F. W. de
 Modulation 36,305
 I. Modulation of a sinusoidal carrier 36,309
 II. Modulation of pulse trains 36,329
 III. Quantization and coding of analog signals 36,337
 IV. Transmission of digital signals 36,343
 V. Modulation in telecommunication 36,353
 —, see Aagaard, E. A.
- Vromans, J. S., see Voorman, J. O.
- Vucht, J. H. N. van
 Intermetallic compounds; background and results of twenty years of research 36,136
 —, see Fransen, J. J. B.
 —, see Verbakel, J. M. M.
- Waalwijk, J. M. and N. Wiedenhof
 The Institute for Nuclear Physics Research 'has finished its work' (35 years of IKO) 39,286
- Waard, P. J. de, see Acket, G. A.
- Waarde, K. M. van der, see Frens, G.
- Wacyk, I. T., see Amato, M.
- Waumans, B. L. A.
 Software aspects of the PHIDIAS system 40,262
- Weaver, J. A., see Saraga, P.
- Weber, C., see Nieuwland, J. M. van
- Weijer, P. van de and R. M. M. Cremers
 Laser diagnostics for low-pressure mercury discharges 43, 62
- Weijters, C. H. L., see Theuwissen, A. J. P.
- Weiss, H., see Dössel, O.
 —, see Klotz, E.
- Welker, T., see Raue, R.
- Werkhoven, C., C. van Oopdorp and A. T. Vink
 Influence of crystal defects on the luminescence of GaP 38, 41
- Wernicke, R., see Daniels, J.
 —, see Hagemann, H.-J.
 —, see Mauczok, R.
 —, see Petersen, A.
- Wesselink, G. A., see Bloem, J.
- Whiffin, P. A. C., see Brice, J. C.
- Whipps, P. W., see Jagt, J. C.
- Wiedenhof, N., see Waalwijk, J. M.
- Wielenga, D. K., see Elst, J. H. R. M.
- Wierenga, P. E. and A. J. J. Franken
 Indentation measurements on thin films 42, 85
- Wijers, J. L. C.
 Three special applications of the Philips high-speed spark-machining equipment 40,199
- Wijn, J. C., see Haisma, J.
- Wijngaarden, H. van, see Galenkamp, H.
- Willems, J. J. G.
 Investigation of a new type of rechargeable battery, the nickel-hydride cell 43, 22
- Willems, L. F., see Hart, J. 't
- Willich, P. and D. Obertop
 Electron-probe microanalysis of thin films 42,162
- Willmott, F. W., see Dolphin, R. J.
- Wit, H. J. de and K. Jager
 Magnetic domains in amorphous alloys for tape-recorder heads 44,101
- Witmer, C. H. M., see Druyvesteyn, W. F.
- Wittekoek, S.
 Optical aspects of the Silicon Repeater 41,268
 —, see Bouwer, A. G.
- Witter, K., see Heitmann, H.
- Wolter, J.
 Research on layered semiconductor structures 43,111
- Wolters, D. R.
 Behaviour of the oxide film in MOS devices 43,330
- Young, B. D., see Guildford, L. H.
- Zaengel, T. and G. A. Lens
 Fast pressure gauge for hot and corrosive gases 39,344
- Zalm, P. C., C. W. T. Bulle-Lieuwma en P. M. J. Marée
 Silicon molecular beam epitaxy on GaP and GaAs 43,154
- Zegers, L. E., see Luyt, B. A. G. van



B. A. G. van Luyt and L. E. Zegers, The Compact Disc Interactive system,
PhilipsTech.Rev.44, No.11/12, 326-333, Nov.1989.

The standard for the CD-I system (Compact Disc Interactive) for consumer applications is an extension of the standard for CD-ROM (Read-Only Memory) for professional applications for computers, which in turn is an extension of the standard for CD-DA (Digital Audio). The CD disc contains images, sound, text, and the associated software in digital form. The information is organized in sectors on the disc, each with its own address and a list of contents. There are two levels of error correction, four quality levels for sound and three quality levels for images. This means that quality can be traded against storage capacity and bit rate when the disc is created. The supplier of interactive programs does this by means of an authoring system. The output from the authoring system is the digital information used in manufacturing the 'mother disc'.

T. G. Spanjer, A. A. van Gorkum and W. M. van Alphen, Electron guns for projection television,
PhilipsTech.Rev.44, No.11/12, 348-356, Nov.1989.

Electron guns of a new design have been made for application in cathode-ray tubes for projection television. These guns contain an impregnated cathode, a selective prefocusing lens and an electromagnetic main lens. The electron spot on the phosphor screen can be very small, even at high beam currents (0.185 mm at 4 mA), mainly because of selective prefocusing. This means that the brightness and resolution will be sufficient for high-definition projection television.

R. Raue, A. T. Vink and T. Welker, Phosphor screens in cathode-ray tubes for projection television,
PhilipsTech.Rev.44, No.11/12, 335-347, Nov.1989.

Phosphor screens in tubes for projection television have to meet some critical requirements, since they operate at much higher electron-excitation densities than the screens in conventional direct-view tubes. Their light output should be only slightly sublinear at high excitation densities and should be stable under prolonged intense electron bombardment. This affects the selection of the phosphors and the screen processing. The preparation of screens with the optimum light output and resolution requires a careful evaluation of their optical properties and degradation behaviour.

E. A. Muijderman, C. D. Roelandse, A. Vetter and P. Schreiber, A diagnostic X-ray tube with spiral-groove bearings,
PhilipsTech.Rev.44, No.11/12, 357-363, Nov.1989.

Spiral-groove bearings can replace the ball bearings in the vacuum of a rotating-anode diagnostic X-ray tube. This increases the maximum load the tube can take and solves a number of other problems, such as limited bearing life, difficulties with anode cooling and noise from the bearings. The spiral-groove bearing has to be lubricated with liquid metal, partly because the vapour pressure of the lubricant should be low, partly because it must conduct the anode current. A gallium alloy has been found to be a successful lubricant. In the new Philips MRC200 diagnostic X-ray tube with spiral-groove bearings more than three times as much energy can be handled during a complete cine investigation as in the SRM 100 tube of the previous generation. This means that an investigation of this type takes only half as long with the new tube, so that there are economic advantages as well as reduced patient stress.