The Radio and Electronic Engineer

The Journal of the Institution of Electronic and Radio Engineers

An Electronic IERE Journal?

ELECTRONIC communication networks have made practical the exchange of information and ideas by means other than paper and the spoken word. Material to be exchanged is entered into a computer store through various inputs and is accessed on-line through local terminals, but such networks have been little explored in the United Kingdom from the users' viewpoint. Research is needed to establish their potential and to assess the problems and costs of using them for different purposes, now and in the future. As acceptability of a system to users largely determines the extent of its use, particular attention needs to be given to psychological and ergonomic aspects of these problems.

The British Library has recently awarded grants to Birmingham University and Loughborough University in order to make possible a programme of research into communication through electronic networks. Birmingham will receive £122,000 to supply hardware and software facilities in support of individual research projects, while Loughborough will receive £134,000 to undertake the first project, which is to prepare an electronic journal and to assess its costs, effectiveness, and impact. The grants will run for three years from 1st November 1980 and in a preliminary period the facilities are being set up and tested and the participants trained.

The first project in this programme of research at Loughborough will investigate some of the problems concerned with setting up and using various types of electronic journals. The concept of the electronic journal is one which involves using a computer to aid the normal procedures whereby an article is written, refereed, edited and published. With the help of suitable software an author may enter a text into a system, and the editor, referees, and ultimately the users, can have access to the text at their computer terminals. The subject of this experimental electronic journal will be 'Computer Human Factors', that is the science and technology of people interacting with computers from either a computing or a human-science viewpoint. Contributors, drawn from UK research establishments and having an active interest in the subject, will each contribute at least one research article and one shorter note in each year of the life of the project, entering the articles on-line to the computer at Birmingham or using other methods including an Optical Character Recognition (OCR) reader at Birmingham and a word processor at Loughborough. The research team will try to assess the value of these different forms of input, including their relative costs, and how far users feel that an electronic journal satisfies their needs as a form of communication.

Initially the experimental journal will be confined to individual, refereed papers, arising from research. However, in order to exploit fully the possibilities of the new medium, the research team at Loughborough will investigate other forms of journal material, such as news letters, annotated abstracts and workshop conferences, and also other aspects of journal preparation, for example co-operative authorship and interaction between authors, editors and referees.

This development, potentially capable of ultimately altering radically practices which have been established for centuries, must of course be seriously considered by all engineers. Most IERE members will be familiar, in some degree, with the presentation of other kinds of information through electronic media. How would *you* like to receive *The Radio and Electronic Engineer* in this way?

F. W. S.

John Powell

The Officers and Council of the Institution learned with great sorrow of the death on January 11th, 1981, of its President, Mr John Powell. He was 56 years of age and leaves a widow, a son and two daughters.

The Institution was represented at the funeral service at St Dunstan's Church, Woking, by the Immediate Past President, Professor W. Gosling, and the Secretary, Air Vice-Marshal S. M. Davidson. The Council of Engineering Institutions was represented by the Chairman, Dr P. A. Allaway.

John Powell was born in Blackpool, Lancashire, in 1924 and educated at Palatine School, Blackpool. He served as a radio officer in the Merchant Navy during the war before joining Cable and Wireless in 1946. Then for eight years he worked overseas in Gibraltar, Salisbury, Rhodesia, Jerusalem, Amman and Nicosia.

Subsequently, he specialized in telegraph switching and automatic error correction at Radio House in London. In 1957 he was seconded to Dollis Hill, the Post Office Research Station, and took part in the UK/Sweden and CANTAT coaxial cable projects before returning as Assistant Manager of the Company's Development Department at Smale House in 1961. In 1963 he returned to submarine cable work and supervised installation and commissioning of systems in the Far East.

He was appointed an Assistant Engineer-in-Chief in 1965. He then served as Group Manager (Special Services) for six months and Assistant Head of Corporate Planning for a year before being appointed Head of Corporate Planning in 1971.

He became Chief Engineer (Record Systems) in 1974, Chief Engineer (Traditional Projects) in 1975 and Chief Engineer (Submarine Systems) in 1977.

In 1978 he became Deputy Engineer-in-Chief, responsible for special studies and supervision of the Engineering Department in supporting business activities. The following year he was appointed Engineer-in-Chief.

John Powell obtained a B.Sc. degree in mathematics and physics in 1955 at London University and specialized in physics for a B.Sc. (Special) in 1956. In 1959 he was awarded a M.Sc. for work in the field of electrical discharges in gases.

Mr Powell was also active in the Territorial Army and Volunteer Reserve. He held a Commission in the Royal Signals from 1958 to 1977, becoming a Lieutenant-Colonel and he was awarded the Territorial Decoration in 1970.

Mr Powell was involved in many aspects of the activities of the Institution. He joined the then British Institution of Radio Engineers as a Student in 1949, qualified as a Graduate member in 1953 and became a Corporate Member in 1959. He was transferred to Fellow in 1965. He served on the Council from 1972–75 and as a Vice-President from 1976–78 and from 1979–80, and was a member of the Executive Committee. He was a member and subsequently Chairman of the Membership Committee between 1966 and 1976. He was the IERE representative on Standing Committee B of the Council of the Engineering Institutions.

He was also a Fellow of the Institution of Electrical Engineers and of the Institute of Physics.

Just eighty days before his death John Powell became the 29th President of the Institution and, following the Annual General Meeting on October 23rd, delivered his Inaugural Address on 'Resource Management: a key to immediate improvement in productivity?' His theme dealt with the human aspects of engineering, and particularly motivation, subjects very close to his heart.

* * *

The following appreciation of John Powell's life and work has been received from Mr R. W. Cannon (Fellow), Managing Director, Public Telecommunications, of Cable and Wireless:

'The unexpected death of John Powell brought back to me a visit which a colleague and I paid to him many years ago when he was convalescing after his first bout of serious illness. We could not then have guessed how successful John would be in fighting and, for many years, overcoming the long term effects of that illness. During these years he maintained a remarkable persistence and devotion to his duties even though he must frequently have felt very unwell. Immediately prior to his appointment as President of the IERE, we all thought he had won his long fight and his health improved quite remarkably, making his death a very unexpected shock.

'John Powell's life was marked by a powerful though well-tempered singleness of purpose. He joined Cable and Wireless as a young telegraph operator and first served overseas on our mobile staff. He decided then that his future lay on the technical side of our business and he began, with great diligence, to equip himself to perform outstandingly in that area.

'Studying for formal qualifications whilst on mobile service in remote areas varies (or used to) from difficult to impossible — and I speak from personal experience. John went as far as he could with his studies overseas and then opted for UK service, eventually receiving his Master's degree.

'Throughout his career he never ceased to encourage his staff to emulate his own efforts. He generated great loyalty amongst them and his firmness towards them was always tempered by a great underlying kindness.

'He became a key figure in the Cable and Wireless entry into the world of submarine coaxial cables and in recent years gained the deep respect of, amongst others, the most senior Japanese professional engineers when he chaired the three-nation technical committee which planned and implemented the laying of the Okinawa–Luzon–Hong Kong cable. In fact he was known and respected throughout the entire world's submarine cable fraternity.

'There are all too few people like John, and we shall miss him sadly.'

The Institution's Bye-laws do not lay down formal procedure for such an unprecedented and sad event as the death of its President. At an emergency meeting of the Executive Committee Professor William Gosling, Immediate Past President, was asked to serve as Acting President for the remainder of what would have been Mr Powell's year of office.

A review of magnetic bubble memories and their applications

K. F. BAKER, M.Sc., C.Eng., M.I.E.R.E.*

SUMMARY

Interest in magnetic bubble memory technology has increased enormously over the past year. The levels of activity and investment have reached a stage where it can be confidently predicted that 1981 will see their widespread use as digital storage elements in production equipment. This paper reviews what bubbles are, how they work and the architecture of present-day chips. The device manufacturing technology is outlined and some projections made for future advancements in storage density, device price and performance.

Applications of bubble memory devices are discussed in terms of the market areas and equipment types for which they are eminently suited. A detailed description of some commercial board level bubble memory systems is given to illustrate the performance achievable as well as design considerations that need to be taken into account in using bubbles. Finally reference is made to end-user equipment and applications that incorporate bubble memories.

* Plessey Microsystems Limited, Water Lane, Towcester, Northants NN12 7JN.

1 Introduction

Magnetic bubble memories are this year going to become practical storage devices that will be used in commercial equipment. They have shown promise for many years and have been widely researched and discussed for more than a decade. Real commercial exploitation is only just beginning and the manufacturers' long-standing claims are now in the 1980s becoming reality. The claims are:

- HIGH RELIABILITY —Solid State Technology
- NON-VOLATILITY —Full Data Retention at Zero Power
- LOW POWER —One Microwatt per Bit
- RUGGEDNESS —No Moving Parts
- LOW COST —10mc/bit at Maturity

What this paper sets out to show is that bubble memories have some attractive unique features, that they can be produced economically and that there is a place in the digital systems market for bubble based products.

To understand their role in systems, it is necessary to describe what bubbles are and how they are made. The importance of understanding their special characteristics is discussed with reference to the methods through which they can be interfaced into digital systems.

2 The Magnetic Bubble Device

The magnetic bubble memory device (m.b.d.) has been extensively discussed in published articles (e.g. Refs. 1–3), so only the simplest description will be given, sufficient to explain the characteristics of the devices used in the systems described and how they are made.

The 'bubbles' are, of course, magnetic domains contained in a thin magnetized film. These bubbles are, in fact, cylindrical through the thickness of the material, and appear circular (hence bubbles) when viewed from the top with the aid of suitable equipment using the Faraday effect.

The bubbles are formed by applying a magnetic field perpendicular to a specially prepared thin film of magnetic material; a magnetic field of the correct strength applied in this manner will cause the naturally occurring serpentine domains of the material to contract until most of them disappear. The remainder take up the cylindrical bubble form, and these bubbles remain stable so long as the magnetic field is maintained within 15% or so. The applied field (termed the 'bias' field) is essential for bubble stability, fields in the range 50–100 Oe being required. If the bias field is increased beyond this, the bubble domains tend to collapse. With even stronger fields the whole layer becomes magnetized in one direction only and the material starts to saturate. The upper and lower values of the bias field, between which a bubble is stable, are known as the margins and these are important in determining the ease of use, stability and reliability of m.b.d.s. The limiting velocity for bubble domains is around 10^3 cm.s^{-1} compared with 10^7 and 10^8 cm.s^{-1} for electrons in silicon and GaAs

respectively. This, coupled with their serial organization and field drive, will leave m.b.d.s as slow to medium speed devices.[4]

Fields of the value required to give stable bubbles can readily be provided by permanent magnets, making stable bubble existence independent of any power source, thus laying the foundations of a non-volatile storage medium. The presence or absence of a bubble at a particular location may be used to denote a data '1' or '0' respectively. This, combined with the fact that bubbles may be moved controllably within the supporting film, forms the basis of a data storage device. The desired basic material is one capable of giving a high density of small bubbles and good production yields; the most universally accepted material in this respect is yttrium iron garnet (YIG) doped with calcium and germanium. Bubble sizes are currently around 2 to 4 μm in diameter and it is expected that it will be possible to produce and work with bubbles less than 2 μm within the next few years. The optimum thickness for the magnetic material is somewhat between 0·8 and 1 times the bubble diameter. For practical purposes the magnetic garnet film is deposited on a non-magnetic garnet substrate, usually gadolinium gallium garnet (GGG). A section showing a bubble domain is shown schematically in Fig. 1. The arrows show the direction of the magnetic field. The region marked + corresponds to magnetization of the garnet in the same direction as the bias field; whilst the bubble regions marked − have fields in the opposite direction.
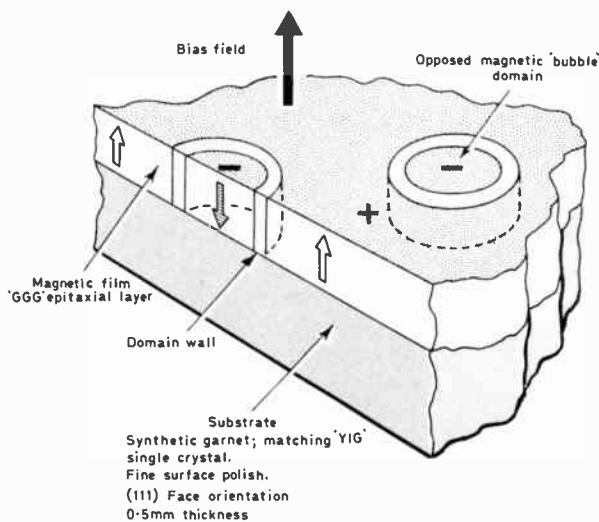

Fig. 1. Bubble domain section schematic.

2.1 Propagation and Detection

Propagation or movement of the bubbles is achieved by varying the localized magnetic field gradients around the bubbles, creating an unbalanced magnetic force on them. The localized magnetic fields are formed by depositing a thin permalloy (soft NiFe alloy) magnetic pattern on to the material containing the bubbles and magnetizing it with an in-plane magnetic field. The stray fields so
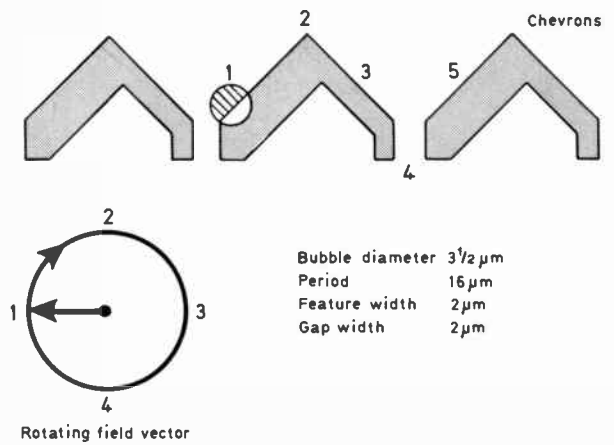

Fig. 2. Bubble movement along chevrons in response to rotating field vector.

produced either repel or attract a bubble, depending on polarity.

By careful choice of pattern it can be arranged that the poles induced by the applied field are very localized, and the bubble positions well defined. Then by rotating this field, the bubbles can be moved along the pattern, one position for each rotation of the field. There are several patterns in current use of which the most effective is the chevron design indicated in Fig. 2.

If, in this conceptual arrangement, we arrange the magnets so that the bubble has a south pole on the surface of the material and the field vector induces a north pole on the left of the chevron, then the bubble resides in position 1 by attraction. If now the vector rotates to position 2, then the north pole shifts to point 2 and the bubble follows, similarly it shifts when the field rotates to 3 and 4. The chevrons are asymmetric to shape the magnetic field lines to ensure that the bubble moves from left to right and not vice versa. The corresponding asymmetry in bubble positions 1, 2, 3 and 4 reflects this field profile. At position 4, the magnetic field lines existing between the two chevrons exert equal and opposite forces on the bubble and this represents a stable or 'rest' position. When the field completes a full rotation back to 1, the bubble transfers to point 5 on the next chevron. So the process repeats itself as the field rotates with the bubble travelling continuously from left to right, for this particular chevron pattern. Turning the chevron structure upside down will cause the bubble to move from right to left for the same field rotation. Special permalloy patterns can be included to constrain the bubble to move around corners for direction reversal. Hence we have an array of permalloy elements that form a magnetic track to guide the bubble motion at a precisely controllable rate determined externally. The rotating field advances the bubble from one stable position to the next for a further rotation of the field. This represents one bit transfer and is the basic element of the bubble shift register.

Data is entered into the shift register by means of a

thin conductor loop (called a hairpin nucleator) placed under an appropriate part of the propagation track. When a current pulse is applied to this loop the local bias field in the garnet film will be modified and bubble generation or annihilation will occur depending on the polarity. A 'write' operation to the memory is performed by generating a new bubble. One way of achieving this is by the division of an existing bubble. With this method the 'seed' bubble remains on a large permalloy generator on the bubble chip. By means of a rotating field it can be stretched until a new bubble breaks off. Control of the generation is by adjustment of the hairpin current amplitude and the rotating field.

Where read-out is non-destructive there is a need for the capability to erase individual bubbles, if the device is to be reprogrammable. This can be achieved by raising the local bias field above the point of bubble collapse by using a current in the opposite direction through the hairpin, or it can be done selectively by directing the bubble on to a large piece of permalloy (the guard ring) where the increased magnetic force will ensure its collapse as the field rotates.

Detection of bubbles is achieved by the magnetoresistive effect[5] which is the change of resistance of a conductor when exposed to a magnetic field. This effect is larger in permalloy than most metals but is still only about 0.15% change for practical bubble fields. Some gain is required to give an adequate output signal. This is achieved by enlarging the bubble field by stretching it to perhaps a hundred bubble diameters by means of a special section of track consisting of lines of chevron elements. These cause the bubble to expand into a strip domain along the chevron stack increasing the stray field of the domain many times. This enlarged bubble passes under an electrically-linked line of permalloy chevrons through which a current of a few milliamperes is passed. The presence of a bubble will cause a resistance change which appears as a voltage change of a few millivolts at the output terminals. The active detector is usually arranged in a bridge network with an identical dummy detector to cancel common mode noise induced by the drive field. This output is fed to a sense amplifier. The drive field rotation is usually restricted to a maximum of 150 kHz because of inductance in the coils and skin effect in the wire. This maximum frequency keeps the drive voltage, current and heat dissipation to reasonable levels.

The organization of a single loop (serial shift register) type device is shown in Fig. 3 on which the key parameters are indicated.

The chevron period of this particular design is 16 μm, with $3\frac{1}{2}$ μm bubbles. Experimental devices are being made with 8 μm periods and $1\frac{3}{4}$ μm bubbles. As the period continues to diminish manufacturing difficulties increase. New techniques (such as electron-beam mask-making and X-ray photolithography) will need to be
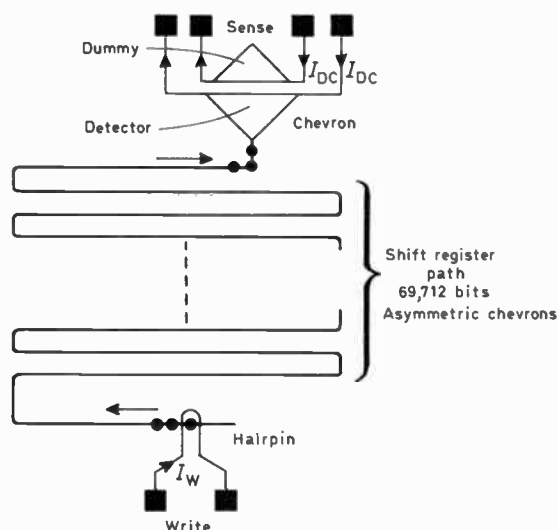

Fig. 3. Shift register organization of 64Kb bubble memory chip.
drive frequency: 100 kHz
Access time: 696 ms max.   power consumption: 300 mW
          348 ms average   package size: $20.3 \times 20.3 \times 8.9$ mm

firmly established before further advances in this area become commercially viable.

The large capacity possible with bubble chips means that for single loop organizations the access time is rather slow and has led to the development of the major-minor or multi-loop organization.

Several other deficiencies of the single-loop device are over-come in the 256 Kbit multi-loop bubble memories which employ the replicate/read block architecture. Instead of a single shift register the structure appears like a large number of short shift registers connected in parallel. These short registers are in the form of closed loops that enable bubbles to recirculate around them. Data can be fed in and out through input and output tracks which connect with all the loops. In this
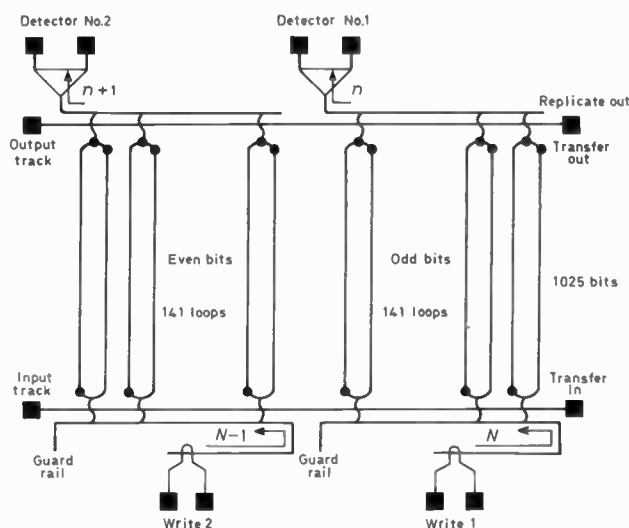

Fig. 4. Block replicate architecture of 256Kb bubble memory chip.
Key Parameters
drive frequency: 150kHz
access time: 8.25 ms (max.)   power consumption: 1 W
          4.85 ms (average)   package size: $30 \times 30 \times 12$ mm

arrangement (Fig. 4) data can be written (and read) on every cycle by splitting the storage blocks into two parts, the loops of one half storing odd bits, whilst those of the other store even bits. Two series-connected nucleators are used, feeding two write lines. The entire data block is therefore generated in each write line, but there is a one-bit difference in path lengths between nucleator and first transfer-in gate for the two halves. This means that when odd bits are aligned for transfer in one half, even bits are similarly aligned in the other half. Bubbles not transferred from each write line are subsequently ejected through to the 'guard rail'. This is the region of permalloy that has an opposing magnetic field which collapses the bubbles. To read, the replicate conductor is activated, odd bits being copied into one read line, and even bits into the other. Path lengths between replicate and detect differ by one bit between each half and allow the data streams to interlace, giving detection on every cycle. Data is read in sequentially via the major write line at the bottom and when this is full a control pulse is applied to the 'transfer-in' conductor which parallel loads the data into the minor loops. At the top end of the minor loop the data may be parallel unloaded into the major read line and propagate sequentially to the detector. The unloading into the read line may be done destructively by 'transferring-out' leaving all zeros in the minor loop. Or by using a different shaped control pulse, data can be 'replicated out' to give non-destructive read-out. The effect on access time is dramatic. No storage location is now more than about a kilobit away from the I/O and the improvement is approximately equal to the number of minor loops.

A 256 K multi-loop memory, having 260 loops of 1025 bits each, would have an average access time to the first bit of 4·85 ms at 150 kHz drive field frequency. This is a considerable improvement on the 348 ms cited in Fig. 3 for a comparable shift-register organized memory. The penalty, however, is the extra peripheral circuitry to drive transfer and replicate gates. The performance specification (see Fig. 4) achieved by such a device is superior in just about every respect when compared with the serial organization. The one drawback is the increased complexity of peripheral circuitry that it requires.

2.2 Redundancy

A single propagation track defect which prevents propagation past it is sufficient to render an entire shift-register chip inoperative, so viable chip yields require very low defect densities in the garnet film and propagation structure. Since the storage in a multi-loop device is distributed among many separate loops, it is possible to employ storage redundancy to increase yields and hence reduce device costs. Extra storage capacity is provided on the chip by having more loops than are actually required in the application for which the device is to be used. For example, the 256 Kbit chip having 260

loops could have, say, 282 loops each with 1025 bits per loop. Any chip could then have up to 22 inoperative loops, yet still meet the storage specification. The chip manufacturer normally specifies that bubbles must not be inserted into defective loops (thus avoiding possible contamination of data stored in good loops) and provides a map of defective loop locations which is generated during device testing. The user must then arrange that zeros are inserted in bit locations in the data block which are destined for defective loops, and also that bits of the block which emanate from defective loops are ignored on reading. Typically, defective loop addresses are stored in a p.r.o.m. which is used with write and read formatting circuits to accomplish the redundant loop skipping.

An alternative in more recently proposed devices such as the TIB 1000 from Texas Instruments, is to store the redundant loop information within the bubble memory device itself. After testing the bubble memory, the manufacturer programs the so-called 'bootstrap loops' with data which indicate which of the 600 loops of the 1 Mbit device are good. The bootloop register in the controller system is then loaded during initialization with this good loop data and this can then decide which loop addresses can be increased for the purpose of reading or writing data. These third generation devices also have built within them additional loops to enable error detection and correction schemes to be implemented easily to minimize data error rates.

3 Bubble Device Manufacture

Bubble device manufacturing technology has evolved over the past few years through developments in materials, circuit layout, photolithography and packaging. Multi-disciplinary teams of research scientists and production technologists have collaborated to achieve what is today the densest form of large scale integration. In many ways this collaboration has benefited from advances in the fabrication of silicon integrated circuits. The use of ion implantation, r.f. sputtering and plasma etching are examples of production techniques familiar to the i.c. technologist. The developments in high resolution mask making and production photoengraving equipment have made possible the very dense circuitry seen in i.c.s. and m.b.d.s today.

Figure 5 gives an overview of the bubble device manufacturing process. It consists of growing a thin garnet film, depositing and photoengraving conductor and permalloy patterns, packaging in a magnetic screen and testing. We will examine each of these stages briefly.

Figure 6 shows a cross-section through part of an m.b.d. chip after the various layers have been deposited and photoengraved. Unlike silicon integrated circuit manufacture there are no diffusion processes and overall the technology is much simpler and has fewer masking
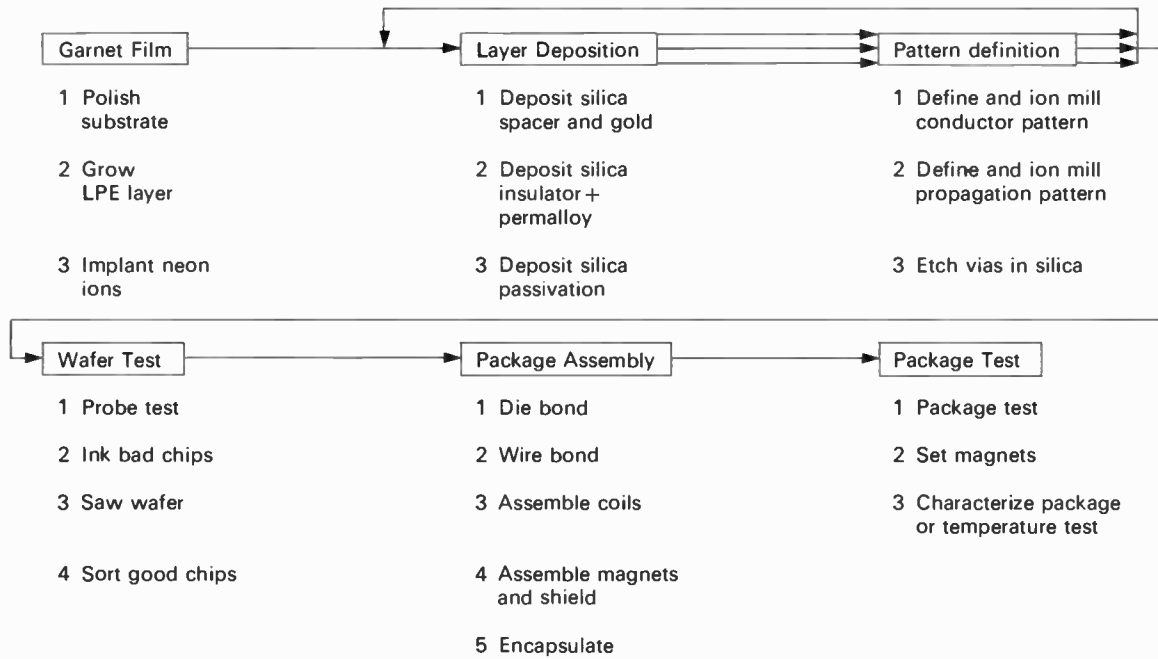
Garnet Film	Layer Deposition	Pattern definition
1 Polish substrate	1 Deposit silica spacer and gold	1 Define and ion mill conductor pattern
2 Grow LPE layer	2 Deposit silica insulator + permalloy	2 Define and ion mill propagation pattern
3 Implant neon ions	3 Deposit silica passivation	3 Etch vias in silica

Wafer Test	Package Assembly	Package Test
1 Probe test	1 Die bond	1 Package test
2 Ink bad chips	2 Wire bond	2 Set magnets
3 Saw wafer	3 Assemble coils	3 Characterize package or temperature test
4 Sort good chips	4 Assemble magnets and shield	
	5 Encapsulate	

Fig. 5. Manufacturing process sequence for bubble memory devices.

stages. The most modern bubble device processes are of the 'planar' type[6] and do not have the large oxide steps indicated in Fig. 6.

The process starts with a 5 cm slice of a non-magnetic garnet cut from a crystal and carefully prepared. The first wafer fabrication stage is to grow the magnetic garnet layer. In this operation a film of GGG
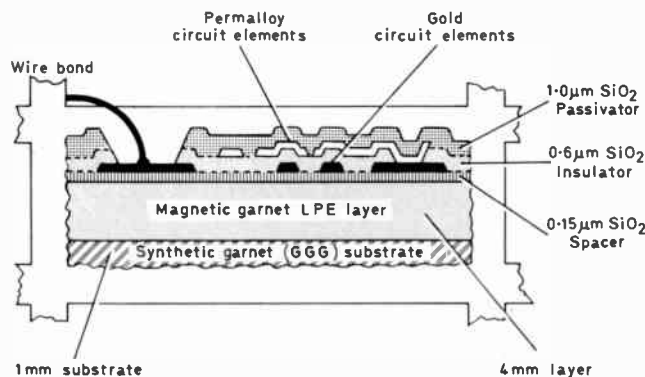

Fig. 6. Cross-section through bubble memory chip showing build-up of layers.

approximately 4 microns thick is grown by liquid phase epitaxy. This simply means dipping the slice into a melt containing the elements of the film under carefully controlled conditions of temperature and time. The control is necessary to ensure that the film is both defect-free and has the required uniaxial anisotropy to form the bubble domains.

The next stage is to deposit a thin layer (0·15 μm) of silicon dioxide over the entire surface of the wafer. This insulating layer helps match the thermal characteristics of the magnetic layer and conductors, is passive in the device operation and is merely there to provide stress relief.

This is followed by the conductor fabrication stage in which gold film conductor is deposited all over the slice on top of the oxide insulating layer. The gold film gives connections to the circuit elements. It is r.f.-sputtered onto the surface of the wafer by using the equipment shown in Fig. 7. The patterns are defined by exposing a photoresist to an ultra-violet light source through a chrome-on-glass mask. To obtain the resolution and edge definition required no wet chemical etching is used; instead the wafer is engraved by ion milling. In this, a collimated beam of high energy argon ions is used to irradiate the entire wafer surface. These are absorbed by areas covered with photoresist but remove the metal that
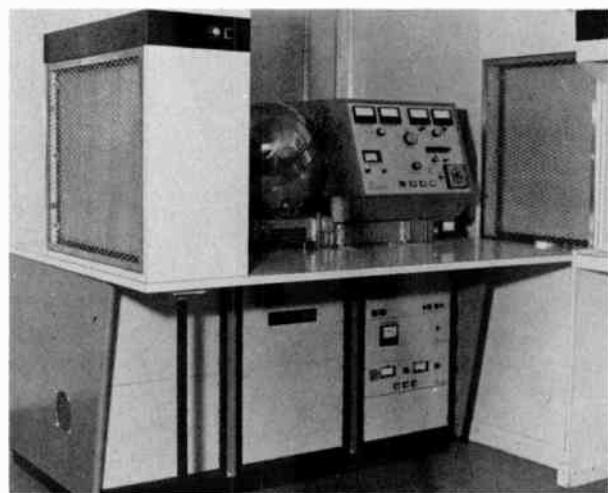

Fig. 7. Equipment used for the sputter deposition of thin metal and insulating layers onto the magnetic bubble slices.

is not covered. The mechanism is simply that the ionic bombardment imparts sufficient energy to the metal to release the bonds holding it on the wafer.

This conductor layer is followed by a 0·6 μm insulator layer of SiO_2 over the entire wafer.

The next stage is the permalloy chevron fabrication. The nickel iron alloy (81% Ni: 19% Fe) is r.f.-sputtered onto the wafer surface and photoengraved to form the chevrons. Again, no chemical etching is used and ion milling removes the unwanted permalloy regions. No close alignment tolerances are required between layers unlike i.c. manufacturing. All the small geometries are defined on the same mask and this reduces photoengraving difficulties.

The wafer is then passivated with a layer of silicon dioxide for protection during the various assembly processes. It is then electrically tested using a computer-controlled probe test machine. Defective chips are marked with an ink spot and the wafer sawn into individual chips using a precision saw. The good chips are then inspected and passed on for packaging. Figure 8 shows the elements of the bubble memory package. It consists of a small p.c.b. onto which the chip is bonded.
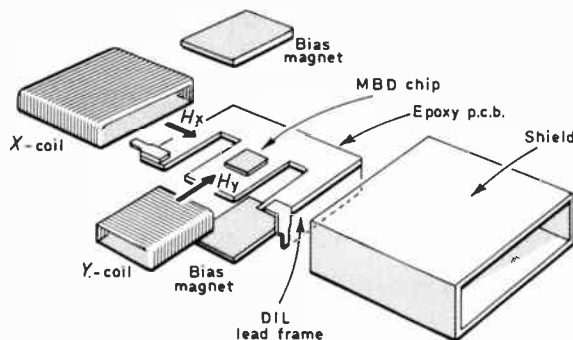

Fig. 8. Package construction for bubble memory assembly.

The gold pads on the chip are then bonded with fine gold wires to tracks on the p.c.b. using i.c.-type bonding equipment. These are then soldered onto a lead frame. The two coils that set up the rotating field are shown in the centre of the figure. They are mounted over the chip orthogonally and permanent magnets are placed above and below them. The whole assembly is then encapsulated within a magnetic screen which forms the body of the package. Final testing of the packaged device then takes place using a computer controlled test station.

The present technology limits packing density to around 10^6 bits per cm^2. Use of electron beam or X-ray lithography may extend this a little[8] but the greater advances leading to 16 M, 64 M and possibly 256 M bit devices will come through structural changes in the circuit elements themselves and the packaging. Several schemes have been proposed[9,10] which illustrate the improvement in packing density that can be achieved, for a given feature size resolved during photolithography. Two possible contenders that have been announced are known as ion-implanted propagation patterns (I2P2) and current-accessed devices.

4 Applications

Bubble systems are not expected to replace disk storage in many commercial applications because they will be too expensive, but for areas where the environment is severe, or 'dusty', or remote, bubbles will prove more reliable and cheaper to maintain. The value of this market is expected to exceed $500 M by 1985. The main market areas have been identified as existing in the telecommunications, machine control, aerospace and defence industries.

The market falls roughly into three segments according to storage capacity (and therefore complexity). These are: (1) small and simple to use systems—often for portable equipment; (2) mid-range systems aimed at cassette/floppy disk replacement type applications; and (3) large systems aimed at rigid disk replacement applications. Products for the first category will probably be custom-built, derived from a basic module design and adapted to a particular customer's board size and interface. The mid-range products in the main will be standard cards which can interface the standard single-board computers. The third category will contain modular systems products that can be interfaced to a range of host computers in a similar way to the disks of today. These will probably be supplied by the traditional computer peripheral companies.

Board or system products fall into three general categories: (1) board-level products that are compatible with existing microcomputer systems, (2) board or boxed sybsystems that can replace disks, and (3) custom-designed systems incorporating bubble memory devices usually in a special configuration.

An example of a board level product is shown schematically in Fig. 9. This is the Plessey PBM 80S, a subsystem consisting of eight bubble memory devices on a single board with the necessary control circuitry. It uses the 64 K serial loop device described previously. The subsystem interfaces the Multibus† in a direct memory-access (DMA) mode implemented with an Intel 8257 programmable DMA controller.

The DMA mode provides fast data transfer (up to 100 Kbytes per second) even though the initial latency is a third of a second. This fast data rate equals the clock rate of the coil drivers. Most of the 20 W power that this board consumes is necessary for the Multibus interface, not the bubbles.

Only 15% of the board area is taken up by bubble devices. The remainder is occupied by the sense and drive circuits, the control and timing. In this design almost half the board area is the DMA and Multibus interface.
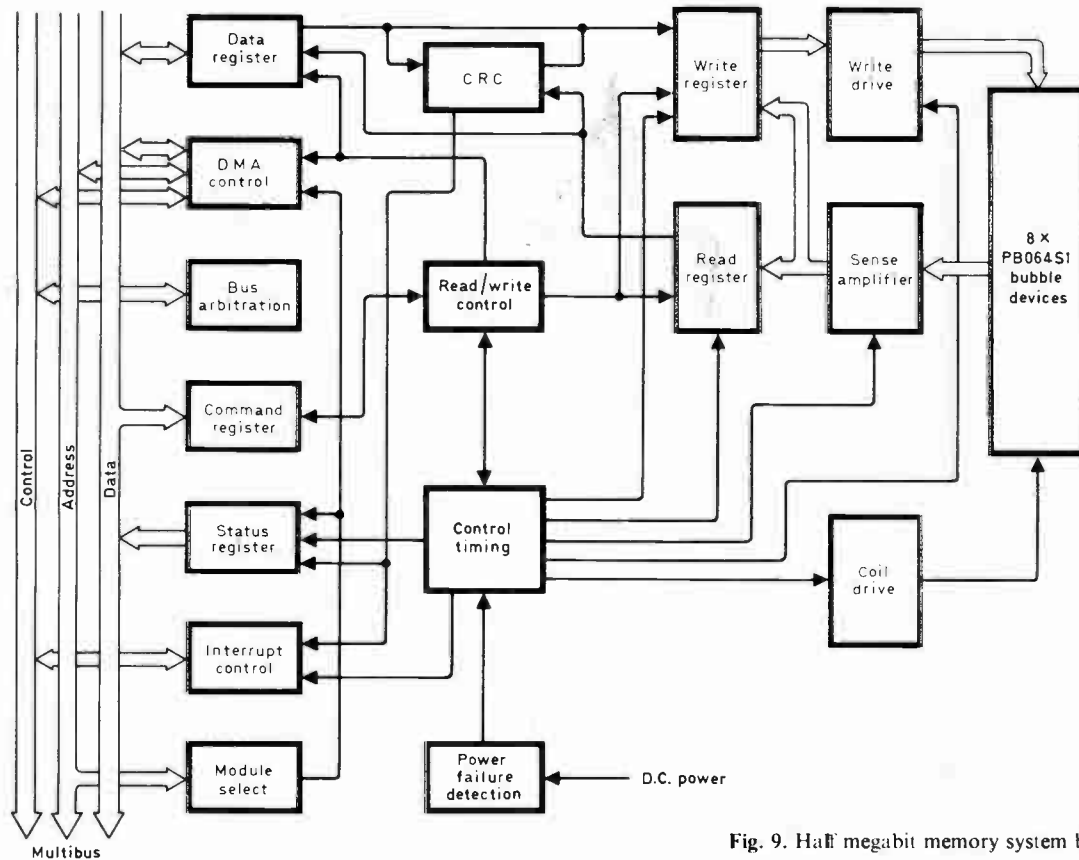
†Multibus is an Intel trade mark.

Fig. 9. Half megabit memory system block diagram.

Bubbles do need a lot of support devices as Fig. 9 shows. Even with a simple serial loop device one needs circuitry to generate timing pulses, drive coils, format and write data and sense outputs. Many currently available m.s.i. logic circuits and several custom circuits have been developed especially to support bubble memory devices.[11-13]

Although bubble memories are rugged under severe environmental conditions, certain design factors must be incorporated in the circuits to prevent damage. For instance, the bubble devices cannot withstand prolonged direct currents, such as those typically used for writing data on to the chip.

Bubble generation requires a current pulse of approximately 200 mA, but this need only be held for approximately 200 ns. This current must pass through the hairpin nucleator which has a resistance of approximately $10\ \Omega$. If a direct current of 200 mA is applied the hairpin element will act like a very fast fuse and the device would be ruined. Therefore, transformer coupled drive circuits are used to stop any d.c. from reaching the generator circuit and a 2% duty cycle is suitable for even the highest data rate. The other sensitive design areas are the triangular wave coil drive circuits and the differential amplifiers to sense the few millivolts from the bubble memory detector outputs. In the PBM 80 S the drive and sense functions are physically very close to the eight bubble memory devices.

This is very necessary to give error-free operation. Most of the data operations on this card are handled by the DMA controller. System control is programmed in p.r.o.m.s which provide all the timing for data transfers as well as the bubble drive timing. On request the DMA chip generates a sequential memory address that permits reading or writing directly to or from the bubbles.

Acquisition of the Multibus for data transfers is accomplished via the bus request function within the controller. It waits for commands from the bubble memory subsystem requesting an output of data. Then the controller claims control of the Multibus and gives the appropriate control signals to enable the bubble memory to read or write directly to or from the system main memory with all the addressing being done by the DMA controller chip. The controller counts the number of transfers as they occur and compares them with the number required in the program instruction. When these two match it supplies a terminating signal to the bubble memory. This indicates that the transfer is complete and the system shuts down to conserve power. Since the bubble memory devices are serial shift register structures, they are destructively read. Thus each read cycle must be followed by a write cycle into the same location. The memory is positioned to the desired data block under software control and the DMA controller is then programmed by the processor to write into the system main memory.

The PBM80S uses a cyclic redundancy check (CRC) to minimize data errors that might occur. The particular structure used on the board generates the CRC code over 128 bytes. A single chip circuit (Fairchild 9401) is used to generate the polynomial $(x^8 + 1)$ function used for CRC. Once the 128 bytes have been written, a signal is generated by the DMA controller to initiate the writing of the check byte.

The operating modes of the bubble memory system are defined by the contents of a command register. The c.p.u. can access the system and determine its state by reading a STATUS register. As well as containing the mode of the system, this also gives indication of any CRC errors. Reading the status register completes the transfer operation.

In addition to providing bubble control and the Multibus interface, the PBM 80S also continuously monitors the $+5$, $+12$ and -5 V supplies. When the voltages rise to 5% below nominal the system is enabled for operation. If they drop below, then the system is disabled into data save. Power-up reset is performed by a hardware or software command.

The counters that control the p.r.o.m. controller are reset upon power-up. This is necessary to prevent intervention either from the bus or software at any time during a bubble memory cycle. Such intervention would cause loss of data. On power-fail the voltage monitor circuit causes the memory to finish its present cycle and stop. The counters not affected by the power-fail signal continue to provide control until the cycle has finished.

At system power-up, the host software must identify the current memory position, that is where it was left when last used. This requires an initializing routine in which a unique identifier is stored in the bubble memory so that the handler can restore the memory to its starting position. The memory array on the bubble memory board has been structured into 538 blocks of 128 bytes each, providing a total capacity of 550, 912 bits of usable memory. The identifier called 'start of memory' occupies an additional 256 bytes. Data are handled in blocks of 128 bytes and formed so that each block is followed by a single byte check sum generated by the CRC. This leaves 54 spare bytes. The identifier is transparent in all but the initializing routine.

Applications for this product exist as a program loader or data logger in industrial controls. It can also be used in remote terminals where off-line storage is not needed or the bubble memory can be off-loaded via a communications line.

In addition to this simple-to-use memory board, many other board level products are just being announced or are in development by a dozen or so suppliers.

An example of a more sophisticated product that has been announced recently is the PBM 80 M system which is an extension of the 80S but uses larger 256 K multi-loop devices. This system consists of an 8085 c.p.u.
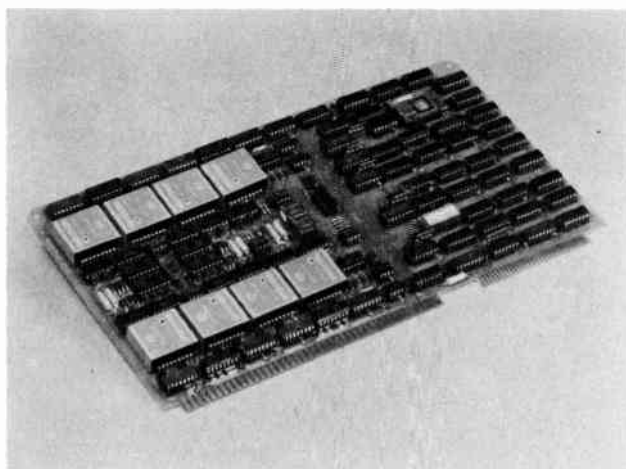

Fig. 10. Two megabit bubble memory board.

memory controller card and up to eight 2Mbit bubble memory cards of the type shown schematically in Fig. 10. These cards fit into a standard System 80 type chassis and provide up to 2Mbytes of storage.

Since the 256 K devices incorporate redundancy, extra firmware and logic are necessary to handle the data formatting and timing. Again DMA and CRC error detection are provided. The complete memory is addressed and all DMA transfers achieved via the controller which interfaces the Multibus and fulfills all the housekeeping supervising and error control functions. The hardware is supported with a software handler package that is incorporated in the monitor program and provides the interface between the user and memory.

Data rates for this system remain fixed at 100 Kbit/s, because of the maximum coil drive frequency, but now the random access time has been reduced by 98% to 6·6 ms through the use of multi-loop devices. Power per bit and cost per bit have also improved because of the greater level of integration. Perhaps more importantly, the resident firmware in the controller p.r.o.m. makes the 80M easier to link into existing application programs.

Military applications for bubble memory abound because the devices are so rugged.[14-16] For example, bubble memories can be set up to replace fixed-head disk (f.h.d.) systems. The PBM 90M for example, is a 1Mbyte module for f.h.d. replacement systems. This module consists of three p.c.b. assemblies containing a total of thirty-two 256 K devices, with their support devices and logic as shown schematically in Fig. 11. This Figure again shows the registers associated with the bus interface, the control timing and redundancy p.r.o.m.s. along with the bubble memory drive and sense circuits.

Six of these modules are connected with a bubble memory controller to form the 6Mbyte bubble memory system shown schematically in Fig. 12.

The design approach adopted is based on using standard storage modules controlled with programmable logic cards. A 16-bit microprocessor acts
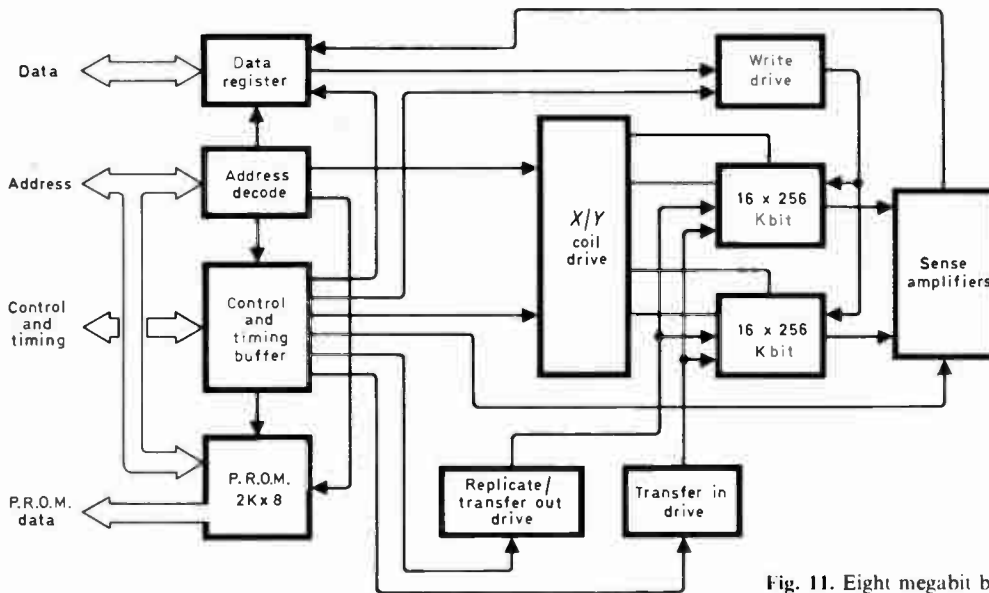
Fig. 11. Eight megabit bubble module block diagram.

as the interface controller between the bubble memory and the host computer. The hardware and command structure of the system are organized so as to permit operation with any 16 bit computer. The store consists of the following components housed in a single compact chassis.

(a) power supply unit

(b) five controller circuit boards comprising:
1. normal interface control board
2. buffer board
3. sector operating control waveform generator board
4. redundancy mapping and CRC control board
5. microprocessor and memory board
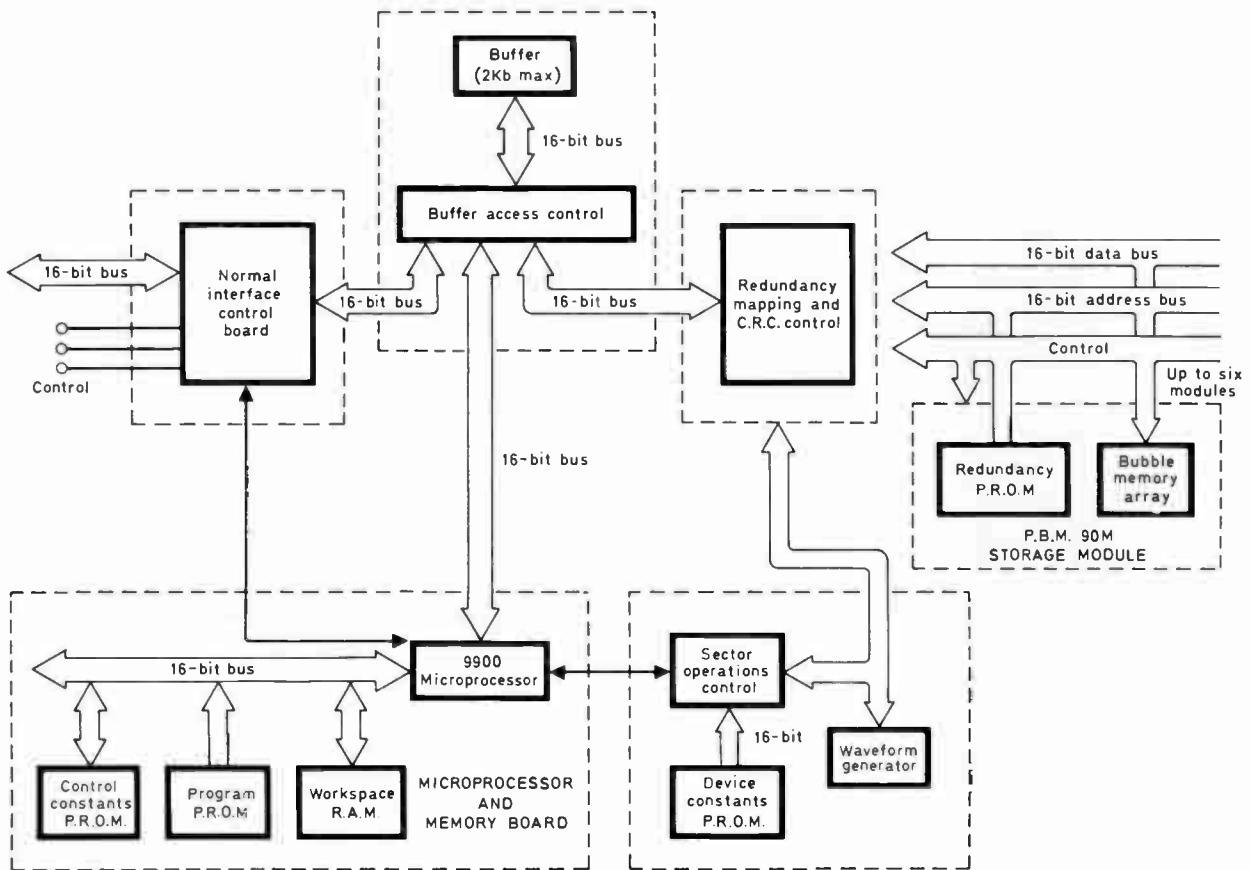
Fig. 12. Six megabyte fixed head disk replacement system block diagram.

(c) six bubble memory modules

The controller is implemented with five modules which enable the system to simulate a fixed head disk through standard disk type commands. These make it possible to read or write into a partial sector, a whole sector or multi-sectors. The memory can respond to processor commands for a break and to acknowledge current status. The seek, restore to zero, initialize, track and check commands that the disk needs for efficient operation are not required by the bubble memory. The operator can perform formatting, self-test, using B.I.T.E. (built-in test equipment) and Clear Down, which erases all the memory at the press of a button.

The specific functions of the controller are as follows:

1. Responds to command bytes from processor to set up transfer parameter bytes.
2. Buffers and monitors data transfers.
3. Generates timing, control and address information to bubble memory modules (PBM90M).
4. Manages bubble memory device redundancy.
5. Manipulates data into and out of bubble devices.
6. Controls internal data formating.
7. Performs CRC error detection.
8. Maintains 'current address' records.
9. Monitors power fails.

Data transfers always take place as 256 word blocks. This makes more than 50Mbits of memory available to the host processor. Average access time to any bit of data is now less than 5ms. The interface to the memory module is organized as 12 groups of 1024 block, each block containing two hundred and fifty-six 16-bit words.

The controller is very flexible and changes can be easily made in four main areas.

(a) Device constants p.r.o.m.—the type of sector operation can be changed by changing this p.r.o.m. The main advantage is that a change of bubble device, to a 1Mbit device, say, would be accommodated by changing this p.r.o.m.

(b) Control constants p.r.o.m.—controller parameters can be varied by altering the contents of this p.r.o.m. This means that the main operating software is not altered.

(c) Host interface control board—the internal interface of this is to the Buffer. All control, data and status information goes over this interface. Thus a change of host processor can be accommodated by replacing this card with another card or cards with a similar internal interface.

(d) Buffer board—this is expandable from two 412 word buffers to two 1024 buffers to accommodate the longer sector size associated with 1Mbit bubble memory devices.

The system is housed, along with the power supplies and B.I.T.E. in the air transport racking (ATR) box shown in Fig. 13. It occupies less than a cubic foot and consumes less than 300W from a 28V d.c. supply. The
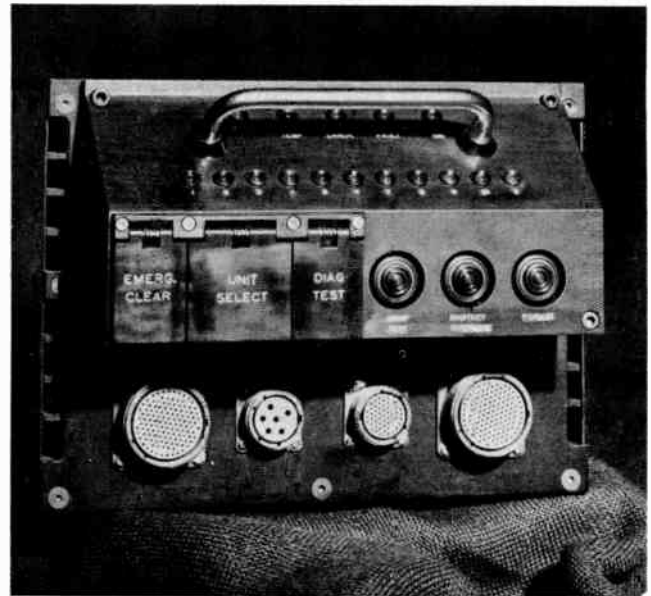

Fig. 13. Six megabyte bubble memory system.

main virtue of this system is its ability to survive in environments that are prone to severe mechanical shock and vibration.

As well as standard memory modules, bubbles are also particularly suitable for customized non-volatile storage because of their small module size. Their ruggedness and lower power consumption make them particularly suitable for portable equipment.

The Plessey Portable Billing Machine[17] is one such system that is now on the market for portable terminals. It is a rugged, lightweight, portable computer developed for immediate billing of electricity accounts. Totally self-contained, it can store all the information, paper and battery power for a day's work between loading operations. Yet it weighs only a few pounds. The system designers have been able to package a bubble memory module, microprocessor logic, a display panel and a keyboard into a single attractive ergonomically designed machine.

The bubble memory that the billing machine incorporates provides the mass storage the application demands. It is mounted on a module with all the drive, sense and control circuitry and yet occupies only 9 cubic inches. No other storage technology could match the size, power, performance, ruggedness, weight and cost attributes of bubbles.

The main advantage of bubble systems is the rugged non-volatile storage medium which has the promise of low cost. They fit into what has been called the access gap between fast, but volatile semiconductor memories and slow, but cheap moving-media memories,[18, 19]

They are expected to appear in most of the main application areas for digital electronics. Their penetration has commenced in telecommunications, data terminals, aerospace and defence applications.

Table 1
Systems reported that incorporate bubble memories

System	Bubble memory capacity	MBD	Application	Manufacturer
13A	272484b	Bell 68K	voice announcement	AT & T/Bell
DSD640	519 KB	TIB0103	disk replacement system	Data Systems Design
765 Terminal	20 KB	TIB0103	portable data capture	T.I.
Dina-mite	100 KB	TIB0203	machine control	Optograms.
GR 7000	40 KB	TIB0203	industrial	G.R. Electronics
TI for USAF	16 Mb	TI 140Kb	airborne	T.I.
Sprint 8T	100 KB	RBM 256	machine control	Siemens
Desk top computer	128 KB	RBM 256	small business computer	Findex
D20 switching system	2·23 Mb	BDG0102	telecommunication	N.T.T. Hitachi
Disk	16 Mb	BDG0102	various	Hitachi
Disk	5 Mb	FBM8031	various	Fujitsu
Programme store	294 Kb	FBM31DB	data terminal	Fujitsu
Terminal	320 Kb	18K	terminal	NEC
Floppy replacement	256 KB	BDG0102	various	OKI
Cassette	8 KB	FBM31DB	data file	Fujitsu
P853 Minicomputer	256 KB	RBM256	traffic control	Philips
Portabubble/81	30K to 120KB	RBM256	portable terminal	Teleram

These markets are expected to expand rapidly over the next few years. As the technology matures and the cost comes down, they are likely to appear in data processing instrumentation and consumer equipment.

Examples of applications will be for small stores in: cash registers, petrol pumps, weighing machines, security systems, machine controllers, amusement machines and vending machines.

Medium-size stores will be used for tape and floppy disk replacement in:
computer terminals, word processing machines, traffic control, PABX's, telex terminals, point of sale terminals, portable data capture systems and ticketing systems.

Large bubble stores will be used as tape or fixed head disk replacement for:
computer mainframes, telephone exchanges, process control, navigation systems and flight management.

Briefly turning back to products that are on the market, there are now five bubble device manufacturers based in the USA and three based in Japan, with many other companies having device manufacturing capability. Device capacities range from 64Kbit to 1Mbit.

There are around twenty board level products that have been announced, mainly intended to be compatible with certain single board computer systems, or aimed at specific applications.

Equipment known to be using bubble memories is listed in Table 1.

For a complete system board at the moment, prices start at about £1000 for a megabit store. One can expect system costs to fall as device prices reduce, and we can expect bubbles to become more competitive with existing technologies over the next five years.

At present bubbles are about the same price as dynamic r.a.m. They are more than twice the price of floppy disk, but have no maintenance cost. They will start to compete with larger disk systems when 1Mbit and greater capacity devices become available in production.

5 Conclusion

Bubble memories have a promising future because they offer high density and non-volatility combined with solid-state technology. No other production manufacturing technique offers all these attributes. The levels of interest and investment throughout the world will ensure adoption of bubble systems in commercial and military applications. The current prices are such that they effectively compete with many media storage systems where ruggedness and reliability are paramount considerations. They offer small module size and are extremely effective for portable systems. The technology advances already announced offer great scope for improved storage, density and performance as well as reduced cost.

6 References

1 Bobeck, A. H., Danylchuk, I., Rossol, F. C. and Strauss, W. 'Evolution of bubble circuits processed by a single mask level', *IEEE Trans. on Magnetics*, **MAG-9**, pp. 474-80, September 1973.
2 Bobeck, A. H., Bonyhard, P. I., and Geusic, J. E., 'Magnetic bubbles—an emerging new memory technology', *Proc. IEEE*, **63**, no. 8, pp. 1176-95, August 1975.
3 Mavity, W. C., 'The RBM256—a block replicate bubble memory', *Electronic Design*, **10**, no. 1, 10th May 1979, pp. 66-70.
4 Vella-Colerio, G. P. and Hess, W. E., 'The generation of rotating magnetic fields for bubble devices', *IEEE Trans. on Magnetics*, **MAG-10**, pp. 750-2, (September 1974)
5 Almasi, G. S., Keefe, G. E., Lin, Y. S., and Thompson, D. A., 'Magneto-resistive detector for bubble domains', *J. Appl. Phys.*, **42**, pp. 1268-9, 1971.
6 Rose, D. K., 'Planar processing for magnetic bubble devices', *IEEE Trans. on Magnetics*, **MAG-12**, pp. 618-21, November 1976.

7   Chang, H., 'Magnetic bubble technology—present and future' (Proceedings of the 7th Conference on Solid-State Devices (Tokyo) *Japan J. Appl. Phys. Suppl.,* 15, pp. 3–10, (1976).

8   Chang, T. H. P., Hatzakis, M., Wilson, A. D., Speth, A. J., Kern, A. and Luhn, H., 'Scanning electron beam lithography for fabrication of magnetic bubble circuits', *IBM J. Res. Dev.,* 20, no. 4, pp. 376–88, July 1976.

9   Juliussen, J. E., 'Magnetic bubble systems approach practical use', *Computer Design,* 15, no. 10, October 1976, pp. 81–91.

10 George, P. K. and Reyling, G., 'Bubble memories come to the boil', *Electronics,* 52, 2nd August 1979, pp. 99–108.

11 Bryson, D., Clover, D., and Lee, D. 'Megabit bubble memory chip gets support from lsi family', *Electronics,* 52, 26th April 1979, pp. 105–111.

12 Bryson, D., Mavity, W. C., et al., 'Designing with bubble memories' Session 27, Digest of WESCON professional program, San Francisco, September 18th–20th 1979.

13 Juliussen, J. E., Lee, D. M., and Cox, G. M., 'Bubbles appearing first as microprocessor mass storage', *Electronics,* 50, 4th August 1977, pp. 81–6.

14 Buvinger, E. A., and Cummings, S. E., 'Military applications of magnetic bubble memory'. Proceedings of the IEEE Electro 77, paper 12.2, New York, 1977.

15 Chen, T. T., Bohning, O. D., Tocci, L. R., Archer, J. L. and Stermer, R. L., 'A magnetic bubble domain flight recorder', *IEEE Trans. on Magnetics,* MAG-10, pp. 739–45, September 1974.

16 Stermer, R. L., 'Bubble memories for spacecraft data recorders', Intermag Digest, paper 29.1, June 1977.

17 Banks, M., 'Portable system prints instant electricity bills', *Computer Weekly,* 26th May 1979.

18 Burford, T. M., 'Applications of serial bubble memory', Intermag Digest, paper 29.2, June 1977.

19 Bonyhard, P. I., Danylchuck, I., Kish, E. E and Smith, J. L., 'Applications of bubble devices', *IEEE Trans. on Magnetics,* MAG-6, pp. 447–50, 1970.

Johnson Matthey wins MacRobert Award

The 1980 MacRobert Award, the premier engineering award in Great Britain, has been won by a team of five men from the Johnson Matthey Group for the development of catalytic systems for controlling pollution from motor vehicle exhaust emissions.

At a private ceremony at Buckingham Palace on 16th December, H.R.H. The Duke of Edinburgh, Senior Fellow of the Fellowship of Engineering, made the presentation of the MacRobert Award and £25,000 to Dr G. J. K. Acres, Dr B. J. Cooper, Mr B. S. Cooper, Dr W. D. J. Evans and Dr D. E. Webster. The Gold Medal was accepted on behalf of Johnson Matthey by the Chairman, Lord Robens of Woldingham.

The MacRobert Award, made annually on behalf of the MacRobert Trusts, is presented in recognition of an outstanding contribution to innovation in the field of engineering, the physical technologies or in the application of the physical sciences, which has enhanced the prestige and prosperity of the United Kingdom of Great Britain and Northern Ireland.

Following increased public concern with the harmful effects of air pollution from motor vehicles in the late 1960s, particularly in the USA and Japan, the Johnson Matthey Group embarked on a major research programme to develop catalytic systems for controlling exhaust emissions.

The US motor industry, faced with regulations requiring a 90% reduction in pollution from cars by carbon monoxide, hydrocarbon and nitrogen oxides, adopted platinum metal catalysts as the most cost-effective method of meeting those regulations. Such catalysts control pollution by oxidising carbon monoxide and hydrocarbons to harmless carbon dioxide and water, while reducing nitrogen oxides to nitrogen. To operate most effectively it is necessary that fuel without added lead be used, but Johnson Matthey is now developing lead-tolerant catalysts for use in Europe.

Johnson Matthey researched, and developed to production, promoted platinum metal catalysts on lightweight ceramic honeycomb supports to meet the motor industry's requirements for a catalyst system capable of low temperature operation.

Production of the catalyst system began in 1974 at Johnson Matthey plants in England and the United States. In 1977, the development was recognized by the granting of the Queen's Award for Technological Achievement to the Johnson Matthey Research Centre.

The Rules and Conditions of the MacRobert Award require the project chosen to have achieved success in the domestic market and likewise to have made a significant contribution to this country's prestige and prosperity in overseas markets. Johnson Matthey technology has been used to advise car manufacturers in the United Kingdom, Europe, America and Japan on the use of catalyst systems for emission control. To date 10 million catalyst units have been supplied to motor companies selling cars in the United States and Japanese markets, including BL Cars, VW-Audi, Rolls-Royce Motors, Aston-Martin-Lagonda, Ford US and General Motors.

Analysis and design of a broadband delta modulation system

N. C. DAVID, M.Eng.*

and

Professor V. MAKIOS, Dipl.Ing., Ph.D., P.Eng.*

SUMMARY

A simple treatment is provided of the analysis of an adaptive delta modulation system for evaluating the performance of the system in terms of complexity, signal-to-noise ratio and dynamic range. The analysis assumes a simple adaptation algorithm and a normal density function input signal with flat spectrum.

A brief description of the implementation of the system is given and finally the experimental results are compared with the theoretical results.

* Laboratory of Electromagnetics, School of Engineering, University of Patras, Patras, Greece.

1 Introduction

Although a great number of contributions to the study of delta modulation systems have been presented,[1-13] only a few of them have attempted to give a satisfactory theoretical analysis of the subject. Most of the contributions refer to computer-based trial and error results[6,7,14] or they are mainly experimental.[8] The purpose of this paper is to derive an equation for the signal-to-noise ratio of any optimum delta modulation system. This equation is then reduced to a simpler form by considering the statistics of the input signal (the input signal is assumed to be a normal density function with flat spectrum) and a simple adaptation algorithm. Through this equation one can directly relate the performance (signal-to-noise ratio) of the system to the variations (dynamic range) of the input signal, and thus reach an optimum for the dynamic range of the system. The verification of the theoretical results in the laboratory confirmed the validity of the theoretical derivations.

2 Mathematical Analysis of a General Adaptive Delta Modulation System and Criterion of Optimum Performance

In this Section the general block diagram of an adaptive delta modulation (a.d.m.) system is presented and a criterion of optimum performance is introduced.

Figure 1 shows the general block diagrams of a complete a.d.m. system: (a) is the coder and (b) the decoder. From this Figure it is apparent that

$$\delta_i = x_i - y_{i-1}, \qquad (1)$$

$$c_i = \delta_i/|\delta_i|, \qquad (2)$$

$$y_i = y_{i-1} + \Delta_i. \qquad (3)$$
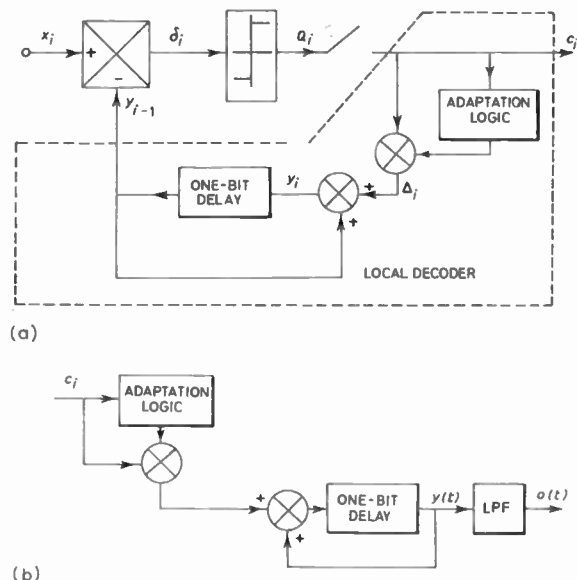

(a)

(b)

Fig. 1. Detailed block diagram of an adaptive delta modulation system.

We define the error (or noise) of the coder as the difference between the input signal x_i and the estimate y_i at the same time instant. Thus:

$$e_i \triangleq x_i - y_i. \tag{4}$$

We also define the signal-to-noise ratio before the low pass filter of the decoder as:

$$\frac{S}{N} \triangleq \frac{E(x_i^2)}{E(e_i^2)} \tag{5}$$

where $E(x_i^2)$ and $E(e_i^2)$ represent the mean square values of x_i and e_i.

We adopt as the criterion of optimum performance of the system the maximization of equation (5).

From equations (1), (3) and (4) we have:

$$\delta_i = x_i - y_{i-1} = x_i - (y_i - \Delta_i) = x_i + \Delta_i - y_i$$
$$= x_i + \Delta_i - (x_i - e_i) = \Delta_i + e_i$$

or

$$e_i = \delta_i - \Delta_i. \tag{6}$$

Combining equations (5) and (6) we get,

$$\frac{S}{N} = \frac{E(x_i^2)}{E[(\delta_i - \Delta_i)^2]}. \tag{7}$$

Unfortunately, no meaningful information can be obtained directly from equation (7) without making some reasonable assumptions relating to the term $E[(\delta_i - \Delta_i)^2]$.

In order to identify reasonable assumptions concerning $E[(\delta_i - \Delta_i)^2]$ we carefully examine the various forms one meets by examining the response of a delta modulator.

The following two cases are being considered:

(a) *Only Overload Error Present.* When the tracking signal is not capable of tracking exactly the input signal $x(t)$ (Fig. 2 region between $t/\tau = 7$ and $t/\tau = 16$) the error introduced by the coder is classified overload error.
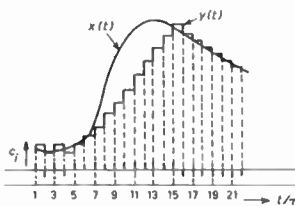

Fig. 2. Response of an a.d.m. coder when quantization error and slope overload error are present.

From Fig. 2 we observe that only in a limited period T (from $t/T = 13$ to $t/T = 16$), in the area of slope overload, the output signal c_i gets more 'ones' than 'zeros' as one would expect for a decreasing input signal $x(t)$ while the rest of the time the output signal gets 'ones' for $x(t)$ increasing and 'zeros' for $x(t)$ decreasing.

(b) *Only Quantization Error Present.* When the signal $y(t)$ tracks exactly the input signal $x(t)$ (Fig. 2 region where $t/\tau < 7$ and $t/\tau > 16$) the error introduced by the coder is called quantization error.

In Fig. 3 the response of an optimum delta modulator is presented. It is assumed that the step size is adjusted in a manner such that at every clock period the mean square error is minimized.
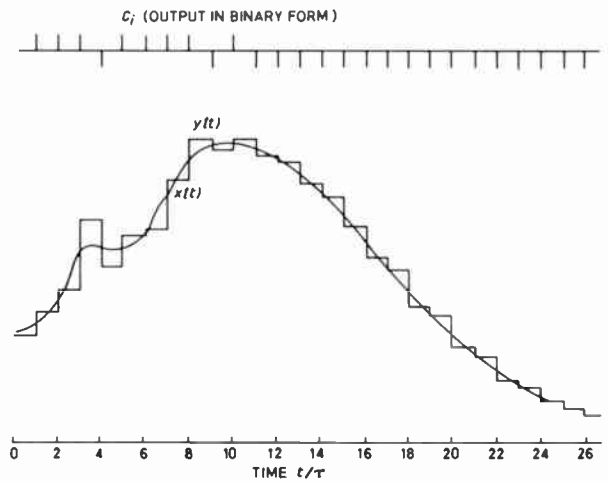

Fig. 3. Optimum response of an a.d.m. coder.

Let us consider one time interval from t_1 to $t_1 + \tau$. The integral squared error between $x(t)$ and $y(t)$ during this period is:

$$G = \int_{t_1}^{t_1+\tau} [x(t) - y(t)]^2 \, dt \tag{8}$$

and $y(t)$ has a constant value h in the range $t_1 < t < t_1 + \tau$. Thus:

$$G = \int_{t_1}^{t_1+\tau} x(t) \, dt - 2h \int_{t_1}^{t_1+\tau} x(t) \, dt + h^2 \int_{t_1}^{t_1+\tau} dt.$$

Differentiating with respect to h we get:

$$\frac{\partial G}{\partial h} = -2 \int_{t_1}^{t_1+\tau} x(t) \, dt + 2h\tau. \tag{9}$$

Setting equation (9) to zero gives:

$$h = \frac{1}{\tau} \int_{t_1}^{t_1+\tau} x(t) \, dt. \tag{10}$$

Considering the next time interval (i.e. from $t_1 + \tau$ to $t_1 + 2\tau$) we may find for the step size h' during this period that:

$$h' = \frac{1}{\tau} \int_{t_1+\tau}^{t_1+2\tau} x(t) \, dt. \tag{11}$$

Subtracting equation (10) from equation (11) we get for the step size at the time under consideration:

$$\Delta = h' - h \frac{1}{\tau} \left[\int_{t_1+\tau}^{t_1+2\tau} x(t) \, dt - \int_{t_1}^{t_1+\tau} x(t) \, dt \right]. \tag{12}$$

In practical delta modulators, the sampling period τ is very small compared to $1/f_b$ (where f_b is the bandwidth of

the input signal). Therefore we may approximate:

$$\int_{t_1}^{t_1+\tau} x(t)\,dt \simeq \tfrac{1}{2}[x(t_1+\tau)+x(t_1)],$$

$$\int_{t_1+\tau}^{t_1+2\tau} x(t)\,dt \simeq \tfrac{1}{2}[x(t_1+2\tau)+x(t_1+\tau)].$$

On the basis of the above approximations equation (12) becomes:

$$\Delta \simeq \tfrac{1}{2}[x(t_1+2\tau)-x(t_1)]. \tag{13}$$

Equation (13) states that, in an optimum delta modulator the step size is positive (negative) for $x(t)$ increasing (decreasing) or equivalently the output of the quantizer is at a high state ($c_1 = 1$) when $dx/dt > 0$ and at a low state ($c_i = -1$) when $dx/dt < 0$. This argument is also confirmed by careful observation of Fig. 3.

In an actual delta modulator the above is not always true because of the overload phenomenon. However, we may say that the probability that $c_1 = 1$ ($c_1 = -1$) is much higher than the probability that $c_i = -1$ ($c_i = 1$) when the derivative of the input signal is positive (negative).

Introducing the variable:

$$\xi(t) \equiv \tau \frac{dx(t)}{dt}$$

(first-order difference of $x(t)$) we can express mathematically the above statement as:

$$P(c_1 = 1) \simeq P(\xi_i > 0),$$
$$P(c_i = -1) \simeq P(\xi_i \leqslant 0). \tag{14}$$

Since c_i is the output of the quantizer (comparator) of the system we may assume that the actual input to the quantizer is ξ_i. Thus we can approximate equation (6) by:

$$e_i \simeq \xi_i - \Delta_i. \tag{15}$$

By virtue of equation (9), equation (7) becomes:

$$\frac{S}{N} = \frac{E(x_i^2)}{E[(\xi_i-\Delta_i)^2]} = \frac{E(x_i^2)}{E(\xi_i^2)}\frac{E(\xi_i^2)}{E[(\xi_i-\Delta_i)^2]}. \tag{16}$$

Finally, an efficient digitizer produces a signal-to-noise ratio which is frequency invariant, i.e. spectrally correlated noise. Thus the ratio of the overall noise to the inband noise is equal to the ratio of half the sampling frequency to the bandwidth of the input signal. For this reason the signal-to-noise ratio of the filtered signal is equal to the unfiltered (S/N) multiplied by $f_s/2f_b$,[16] or,

$$\left(\frac{S}{N}\right)_{LP} = \frac{E(x_i^2)}{E(\xi_i^2)}\frac{E(\xi_i^2)}{E[(\xi_i-\Delta_i)^2]} \cdot \frac{f_s}{2f_b}. \tag{17}$$

Equation (17) does not consider the statistics of the input signal nor the adaptation logic (algorithm). For this reason it may apply to any 'optimum' delta modulation system. In the following Section we consider

the statistics of the input signal and a simple adaptation algorithm.

3 Analysis of an A.D.M. System for a Normal Density Function Input Signal with Flat Spectrum

A delta modulator is a non-linear system and consequently its behaviour is strongly dependent on the input signal. In this Section we derive an expression for the signal-to-noise ratio for a specific input process and a simple adaptation algorithm. We also examine the dependence of the dynamic range on the number of the bit delay of the output signal.

The input signal function of the system is assumed to be normal and stationary with mean value not necessarily zero, a variance S_0 and a flat spectrum given by:

$$S(\omega) = \frac{S_0\pi}{\omega_b} \quad \text{for } \omega \leqslant \omega_b$$

$$S(\omega) = 0 \qquad \omega > \omega_b$$

where ω_b is the bandwidth of the input signal.

Apparently the autocorrelation of this signal is the inverse Fourier transform of $S(\omega)$ given by:

$$R(\tau') = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\exp{(j\omega\tau')}\,d\omega = S_0\frac{\sin\omega_b\tau'}{\omega_b\tau'}. \tag{18}$$

Both $S(\omega)$ and $R(\tau')$ are shown in Fig. 4.

Fig. 4. Power spectrum and autocorrelation of the input signal $x(t)$.

The autocorrelation of $\xi(t)$ is given by:

$$R_\xi(\tau') = E\left[\tau\frac{dx(t)}{dt} \cdot \tau\frac{dx(t+\tau')}{dt}\right] = \tau^2 R_{x'x'}(\tau')$$

where

$$R_{x'x'}(\tau') = E\left[\frac{dx(t)}{dt}\frac{dx(t+\tau')}{dt}\right] = -\frac{d^2R(\tau')}{d\tau'^2}$$

is the autocorrelation of the derivative of $x(t)$. It can be shown (see Appendix) that:

$$\zeta(\tau') \equiv \frac{R_\xi(\tau')}{R_\xi(0)}$$

$$= 3\left[\frac{2\cos\omega_b\tau'}{(\omega_b\tau')^2} + \frac{\sin\omega_b\tau'}{(\omega_b\tau')} - \frac{2\sin(\omega_b\tau')}{(\omega_b\tau')^3}\right]$$

where $\zeta(\tau')$ is the correlation coefficient of $\xi(t)$ and

$$R_\xi(0) \equiv E(\xi_i^2) = \frac{4\pi^2}{3}\left(\frac{f_b}{f_s}\right)^2 S_0 = \lambda^2 S_0 = \lambda^2 E(x_i^2)$$

119

$$\frac{E(\xi_i^2)}{E(x_i^2)} = \lambda^2 \qquad (19)$$

and also:[15]

$$E(c_i) = 0$$

$$R_C(\tau') = E[c(t+\tau')c(t)] = \frac{2}{\pi}\sin^{-1}\zeta(\tau') \qquad (20)$$

where $R_C(\tau')$ is the autocorrelation of the output signal c_i,

$$R_{\xi C}(\tau') = E[\xi(t+\tau')c(t)] = \zeta(\tau')\sqrt{\frac{2R_\xi(0)}{\pi}} \qquad (21)$$

where $R_{\xi C}(\tau')$ is the cross-correlation between ξ and c, or,

$$R_{\xi C}(\tau') = \zeta(\tau')\lambda\sqrt{\frac{2}{\pi}S_0}. \qquad (22)$$

In conclusion we get the following important results for the correlation function of the output signal and the correlation between input and output:

$$c_{ik} = E(c_i c_k) = \frac{2}{\pi}\sin^{-1}\zeta(|i-k|\tau)$$

$$R_{ik} = E(\xi_i c_k) = \zeta(|i-k|\tau)\lambda\sqrt{\frac{2}{\pi}}\sqrt{S_0} \qquad (23)$$

where τ is the sampling period.

In Tables 1(a), (b) and (c) the values of λ, c_{ik} and $\zeta(\tau)$

Table 1

Values of c_{ik} and $\zeta(\tau)$ for three different sampling rates as calculated from equations (19) and (23). $n = |i-k|$

(a) $(f_s/f_b) = 8$; $\lambda = 0.4534$

n	$\zeta(n\tau')$	$c_{ik} = c_n$
1	0.822	0.614
2	0.362	0.236
3	-0.188	-0.121
4	-0.068	-0.416

(b) $(f_s/f_b) = 12$; $\lambda = 0.30229$

n	$\zeta(n\tau')$	$c_{ik} = c_n$
1	0.919	0.742
2	0.692	0.486
3	0.362	0.236
4	-0.009	-0.0057

(c) $(f_s/f_b) = 16$; $\lambda = 0.2267$

n	$\zeta(n\tau')$	$c_{ik} = c_n$
1	0.954	0.806
2	0.822	0.614
3	0.617	0.423
4	0.362	0.236

are tabulated for three different sampling rates as they were calculated from equations (23) and (19).

At this point we introduce the adaptation algorithm. We assume that:

$$\Delta_i = \sum_{k=0}^{n-1} \alpha_{n-k} c_{i-k} \qquad (24)$$

where α_{n-k}, $k = 0, 1, 2, \ldots, n-1$ are constants with $\alpha_n \neq 0$.

Equation (24) illustrates the dependence of Δ_i on the present and $(n-1)$ previous states of the output binary signal c_i (n-bit dependence).

Expressing $E[(\xi_i - \Delta_i)^2]$ in terms of equation (24), we have:

$$E[(\xi_i - \Delta_i)^2]$$

$$= E\left[\left(\xi_i - \sum_{k=0}^{n-1}\alpha_{n-k}c_{i-k}\right)^2\right]$$

$$= E(\xi_i^2) - 2\sum_{k=0}^{n-1}\alpha_{n-k}E(\xi_i c_{i-k}) + E\left[\left(\sum_{k=0}^{n-1}\alpha_{n-k}c_{i-k}\right)^2\right]$$

$$= E(\xi_i^2) - 2\sum_{k=0}^{n-1}\alpha_{n-k}E(\xi_i c_{i-k}) + \sum_{k=0}^{n-1}\alpha_{n-k}^2 E(c_{i-k}^2) + $$

$$+ \sum_{k=0}^{n-1}\alpha_{n-k}\sum_{\substack{l=0\\l\neq k}}^{n-1}\alpha_{n-l}E(c_{i-k}c_{i-l}).$$

Since $E(c_{i-k}^2) = 1$ and $E(\xi_i c_{i-k})$ and $E(c_{i-k}c_{i-l})$ are given by equations (23) the above takes the form:

$$E[(\xi_i - \Delta_i)^2]$$

$$= R_\xi(0) - 2\lambda\sqrt{\frac{2}{\pi}}\sqrt{S_0}\sum_{k=0}^{n-1}\alpha_{n-k}\zeta(k\tau) + $$

$$+ \sum_{k=0}^{n-1}\alpha_{n-k}^2 + \sum_{k=0}^{n-1}\alpha_{n-k}\sum_{\substack{l=0\\l\neq k}}^{n-1}\alpha_{n-l}c_{kl} \qquad (25)$$

$$= \lambda^2 S_0 - 2\sqrt{\frac{2}{\pi}}\sqrt{S_0}\sum_{k=0}^{n-1}\alpha_{n-k}\zeta(k\tau) + $$

$$+ \sum_{k=0}^{n-1}\alpha_{n-k}^2 + \sum_{k=0}^{n-1}\alpha_{n-k}\sum_{\substack{l=0\\l\neq k}}^{n-1}\alpha_{n-l}c_{kl}.$$

Denoting

$$\left.\begin{array}{l}\mu = \lambda\sqrt{\frac{2}{\pi}}\sum_{k=0}^{n-1}\alpha_{n-k}\zeta(k\tau) \\[2mm] m^2 = E(\Delta_i^2) = \sum_{k=0}^{n-1}\alpha_{n-k}^2 + \sum_{k=0}^{n-1}\alpha_{n-k}\sum_{\substack{l=0\\l\neq k}}^{n-1}\alpha_{n-l}c_{kl},\end{array}\right\} \quad (26)$$

equation (16) becomes:

$$\frac{S}{N} = \frac{1}{\lambda^2}\frac{\lambda^2 x^2}{\lambda^2 x^2 - 2\mu x + m^2} = \frac{x^2}{\lambda^2 x^2 - 2\mu x + m^2} \qquad (27)$$

where x is the root mean square value of the input signal. Obviously, equation (27) should be positive or,

$$\lambda^2 x^2 - 2\mu x + m^2 \geqslant 0.$$

Equation (27) is always true if the inequality condition accepts only imaginary solutions, i.e.

$$\mu^2 - \lambda^2 m^2 \leqslant 0. \tag{28}$$

Introducing

$$\gamma^2 \equiv \frac{\lambda^2 m^2}{\mu^2} \tag{29}$$

we obtain

$$\gamma^2 \geqslant 1. \tag{30}$$

It is reasonable to assume that

$$\alpha_n > \alpha_{n-1} > \alpha_{n-2}, \ldots, > \alpha_1 \geqslant 0. \tag{31}$$

Besides, we want the step size to be positive for $c_i = 1$ and negative for $c_i = -1$. In other words, the following condition should be valid:

$$\alpha_n \geqslant \sum_{k=1}^{n-1} \alpha_{n-k} > 0. \tag{32}$$

Assuming λ, m and μ are fixed, the maximum value of S/N is found with the aid of equation (27):

$$\frac{d(S/N)}{dx} = 0$$

or

$$x_{max} = \frac{m^2}{\mu}$$

where x_{max} is the r.m.s. value of the input signal for which the S/N becomes maximum assuming fixed values for m, λ and μ. Then,

$$(S/N)_{max} = \frac{m^2}{\lambda^2 m^2 - \mu^2} \tag{33}$$

or

$$(S/N)_{max} = \frac{1}{\lambda^2} \frac{\gamma^2}{\gamma^2 - 1}$$

$$(S/N)_{max} = 10 \log \left(\frac{1}{\lambda^2} \frac{\gamma^2}{\gamma^2 - 1} \right) \quad \text{dB.} \tag{34}$$

Another interesting parameter is the dynamic range. We define the dynamic range as:

$$DR = 20 \log (x_1/x_2) \tag{35}$$

where x_1 and x_2 are the r.m.s. values of the input signal for which the S/N falls by 3 dB relative to the peak value. That is:

$$x_1 = \frac{m^2}{\mu} \frac{1 + \sqrt{\gamma^2 - 1}}{2 - \gamma^2},$$
$$x_2 = \frac{m^2}{\mu} \frac{1 - \sqrt{\gamma^2 - 1}}{2 - \gamma^2}. \tag{36}$$

We want x_1 and x_2 to be both positive. Thus:

$$2 - \gamma^2 \geqslant 0$$

or

$$\gamma^2 \leqslant 2$$

and with the condition (30) we obtain:

$$1 \leqslant \gamma^2 \leqslant 2 \tag{37}$$

and

$$DR = 20 \log (x_1/x_2) = 20 \log \left(\frac{1 + \sqrt{\gamma^2 - 1}}{1 - \sqrt{\gamma^2 - 1}} \right). \tag{38}$$

From equations (35) and (38) the following useful conclusions may be deduced:

(a) The $(S/N)_{max}$ becomes greater as $\gamma^2 \to 1$. However, the dynamic range increases as $\gamma^2 \to 2$. This indicates that one cannot simultaneously get an optimum $(S/N)_{max}$ and a wide dynamic range. Thus γ^2 should be selected in a fashion that both the $(S/N)_{max}$ is acceptable and the dynamic range is wide.

(b) Another important conclusion is that γ^2 affects the dynamic range more than it affects the $(S/N)_{max}$ (as $\gamma^2 \to 2$).

The second conclusion becomes more apparent by considering an example. For a dynamic range of 40 dB we have:

$$20 \log \frac{1 + \sqrt{\gamma^2 - 1}}{1 - \sqrt{\gamma^2 - 1}} = 40$$

or

$$\gamma^2 = 1.96$$

and thus,

$$(S/N)_{max} = 3.11 - 10 \log \lambda^2.$$

For a dynamic range of 20 dB we have:

$$20 \log \frac{1 + \sqrt{\gamma^2 - 1}}{1 - \sqrt{\gamma^2 - 1}} = 20$$

or

$$\gamma^2 = 1.669$$

and thus,

$$(S/N)_{max} = 3.96 - 10 \log \lambda^2.$$

It becomes apparent that for a gain of 20 dB in the dynamic range one loses only $3.93 - 3.11 = 0.85$ dB in the $(S/N)_{max}$ value. In the following Section the optimization of γ^2 in terms of the dynamic range and $(S/N)_{max}$ is examined.

4 Dependence of γ^2 on the Number of the Bit Delay of the Output Signal

Substituting the values of m and μ in equation (29), we get the following expression for γ^2:

$$\gamma^2 = \frac{\sum\limits_{k=0}^{n-1} \alpha_{n-k}^2 + \sum\limits_{k=0}^{n-1} \alpha_{n-k} \sum\limits_{\substack{l=0 \\ l \neq k}}^{n-1} \alpha_{n-l} c_{kl}}{\left[\sum\limits_{k=0}^{n-1} \alpha_{n-k} \zeta(k\tau) \right]^2} \frac{\pi}{2}.$$

Introducing

$$b^2 \equiv \frac{2}{\pi} \gamma^2 \quad \text{with} \quad \frac{2}{\pi} \leqslant b^2 \leqslant \frac{4}{\pi}$$

we get

$$b^2 = \frac{\sum_{k=0}^{n-1}\alpha_{n-k}^2 + \sum_{k=0}^{n-1}\alpha_{n-k}\sum_{\substack{l=0\\l\neq k}}^{n-1}\alpha_{n-k}c_{kl}}{\left(\sum_{k=0}^{n-1}\alpha_{n-k}\zeta(k\tau)\right)^2}. \qquad (39)$$

The variations of b^2 as a function of the number of bit delay (n) will be studied.

1. Linear Delta Modulation $(n = 1)$

In this case equation (39) gives:

$$b^2 = 1.$$

This indicates that the dynamic range of a linear delta modulation system is independent of the step size and the sampling frequency. This is in complete agreement with the results obtained by other investigators[6] employed by computer-based trial and error procedures (see Fig. 5).
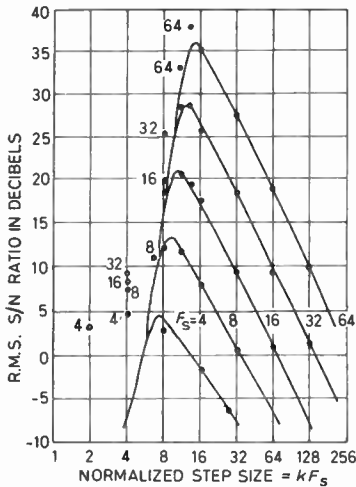
Fig. 5. Signal/noise ratio for bandlimited flat Gaussian signals. (After O'Neal.[6])

For maximum S/N the step size can be specified from the relationship:

$$\dot{x}_{max} = \frac{m^2}{\mu}.$$

Since (see equations (26)),

$$m^2 = E(\Delta_i^2) = \Delta^2$$

$$\mu = \lambda\sqrt{\frac{2}{\pi}}\Delta$$

we obtain,

$$x_{max} = \frac{\Delta^2}{\lambda\sqrt{\frac{2}{\pi}}\Delta}$$

or

$$\Delta = \lambda\sqrt{\frac{2}{\pi}}x_{max}.$$

The dynamic range is $\left(\gamma^2 = \frac{\pi}{2}b^2 = \frac{\pi}{2}\right)$

$$DR = 20\log\frac{1+\sqrt{\left(\frac{\pi}{2}\right)^2-1}}{1-\sqrt{\left(\frac{\pi}{2}\right)^2-1}} = 17\cdot12\ dB$$

and

$$(S/N)_{max} = 4\cdot39 - 10\log\lambda^2.$$

2. Two-bit Delay A.D.M. $(n = 2)$

In this case equation (39) gives:

$$b^2 = \frac{\alpha_2^2+\alpha_1^2+2\alpha_1\alpha_2c_{12}}{\alpha_2^2+\alpha_1^2(\tau)+2\alpha_1\alpha_2\zeta(\tau)} = \frac{1+Q^2+2Qc_{12}}{1+\zeta^2(\tau)+2\zeta(\tau)Q}$$

(note $c_{12} = c_{21}$), where $Q = \frac{\alpha_1}{\alpha_2}$ and $0 \leqslant Q \leqslant 1$.

Figure 6 plots b^2 vs. Q for the three sampling rates of interest. From this Figure we observe that b^2 does not exceed 1, i.e. the dynamic range for $n = 2$ is less than or equal to the dynamic range for $n = 1$. We also observe that $b^2 = 1$ when $Q = 0$ or $\alpha_1 = 0$. This is of course the case of a linear delta modulator $(n = 1)$.
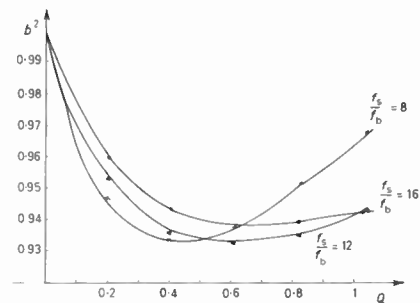

Fig. 6. Variation of b^2 vs. Q for two-bit delay $(n = 2)$.

3. A.D.M. for $n \geqslant 3$

In this case we relate the parameters α_{n-k}, $k = 0, 1, 2, \ldots, n-1$ as follows:

$$\alpha_{n-k} = \alpha Q^{f(k)} \qquad (40)$$

where $\alpha > 0$, and $0 < Q \leqslant 1$, $f(k)$ is a function increasing monotonically with k and α is a constant to be specified.

Since $0 < Q \leqslant 1$ and $f(k)$ is increasing with k, the constraint indicated by (31) is satisfied.

The constraint given by (32) is satisfied if

$$\alpha Q^{f(0)} \geqslant \alpha\sum_{k=1}^{n-1}Q^{f(k)}$$

or

$$Q^{f(0)} \geqslant \sum_{k=1}^{n-1}Q^{f(k)}. \qquad (41)$$

Now the problem has been reduced to determining the proper values of Q and $f(k)$ such that equation (32) is satisfied and b^2 obtains a value as close as possible to $4/\pi = 1.2732$, a value for which we get an infinite dynamic range (in theory).

In order to compute the dynamic range, it is necessary to define appropriate values of Q and $f(k)$. To this point we have assumed:

$$0 < Q \leqslant 1,$$

$$f(k_1) \geqslant f(k_2), \quad \text{if } k_1 \geqslant k_2,$$

$$Q^{f(0)} \geqslant \sum_{k=1}^{n-1} Q^{f(k)}.$$

In order to compute values for the dynamic range, we assume a number of specific analytic expressions for $f(k)$.

Initially we assume:

$$f(k) = pk, \quad p \geqslant 0. \tag{42}$$

Then

$$1 \geqslant \sum_{k=1}^{n-1} Q^{kp}$$

or

$$\sum_{k=0}^{n-1} Q^{kp} \leqslant 2$$

or

$$\frac{1 - Q^{np}}{1 - Q^p} \leqslant 2.$$

Figures 7, 8 and 9 illustrate how b^2 varies with respect to p for the three sampling rates of interest and for three, four and five-bit delay respectively.

Careful observation of these Figures leads to the following conclusions:

When the sampling rate is small ($f_s/f_b = 8$), b^2 exceeds 1, that is, the dynamic range is improved as compared to the dynamic range obtained by linear delta modulation ($b^2 = 1$).

Also as n increases, the maximum value of b^2 increases or equivalently the dynamic range increases.

Thus, for $f_s/f_b = 8$ and $n = 3$, $b^2_{max} = 1.0152$; for $n = 4$, $b^2_{max} = 1.07433$; for $n = 5$, $b^2_{max} = 1.12195$.

Therefore, we are confident that increasing the value of n to $n = 6, 7, 8, \ldots$ we shall finally get a value for b^2 close to the desired value $4/\pi$.

Another interesting observation from the graphs is that at higher sampling rates the values of b^2_{max} are smaller for a given n. This means that higher sampling rates are more demanding in bit delay for a wide dynamic range.

If we continue trying larger values of n, we may finally find a satisfactory b^2. However, the larger the value of n the more complicated the implementation of the circuit becomes. For this reason, we consider another function $f(k)$, namely:

$$f(k) = k^p, \quad p \geqslant 0. \tag{43}$$

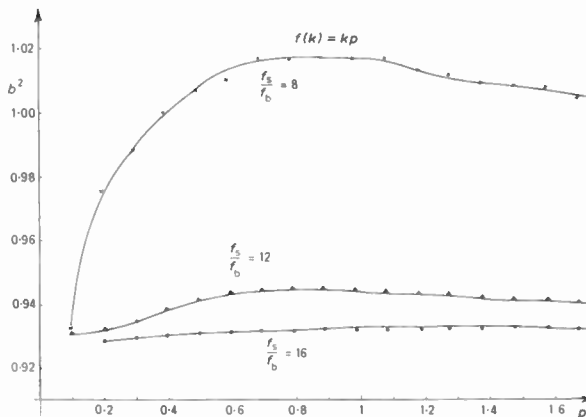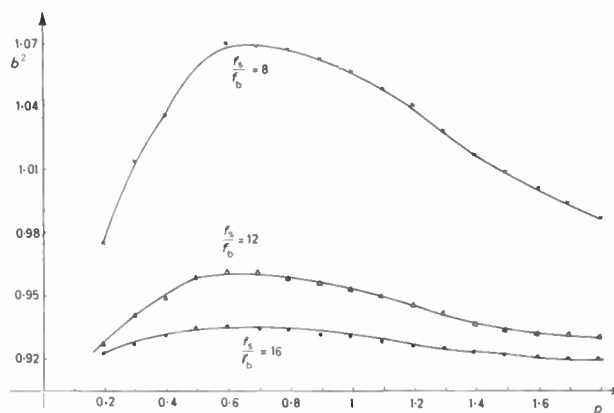Fig. 7. Variation of b^2 vs. p for three-bit delay ($n = 3$) and $f(k) = kp$.

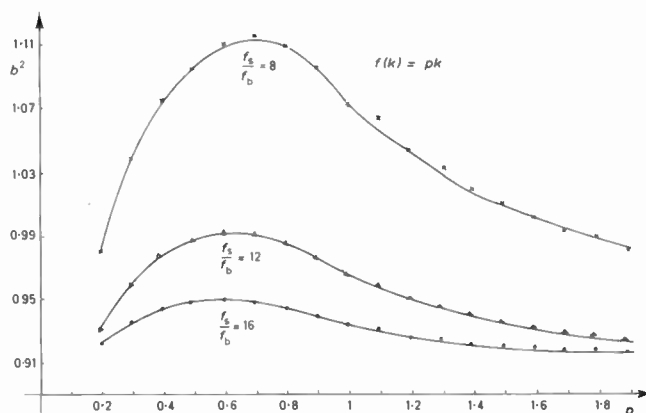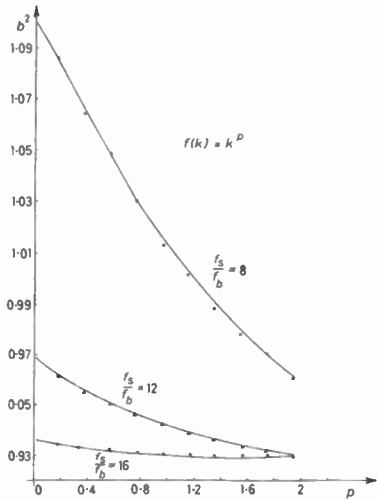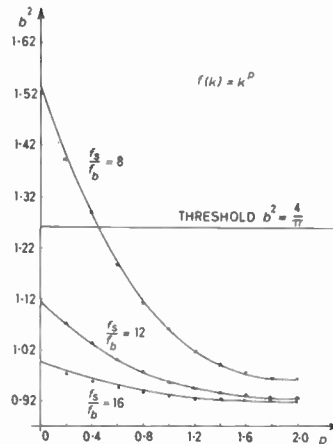Fig. 8. Variation of b^2 vs. p for four-bit delay ($n = 4$) and $f(k) = kp$.

Fig. 9. Variation of b^2 vs. p for five-bit delay ($n = 5$) and $f(k) = kp$.

In this case,

$$Q^0 \geqslant \sum_{k=1}^{n-1} Q^{k^p}$$

or

$$\sum_{k=0}^{n-1} Q^{k^p} \leqslant 2.$$

Fig. 10. Variation of b^2 vs. p for three-bit delay ($n = 3$) and $f(k) = k^p$.

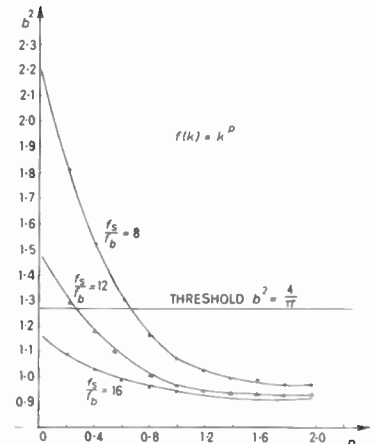**Fig. 11.** Variation of b^2 vs. p for four-bit delay ($n = 4$) and $f(k) = k^p$.

Fig. 12. Variation of b^2 vs. p for five-bit delay ($n = 5$) and $f(k) = k^p$.

We again plot b^2 vs. p for the three different sampling rates of interest and $n = 3, 4, 5$ as shown in Figs 10, 11 and 12.

The conclusions to be drawn from these Figures are similar to those for the case $f(k) = kp$. However, we observe that in the present case ($f(k) = k^p$), b^2 approaches the desired value $4/\pi$ for a wider number of sampling rates. We also observe that for $n = 4$ and $f_s/f_b = 8$, b^2 reaches its boundary value $4/\pi$. This means that any further increase in n (bit delay) would not improve the dynamic range for this particular sampling rate.

Similarly, from Fig. 12 we see that for $n = 5$, b^2 reaches its boundary value ($4/\pi$) for both $f_s/f_b = 8$ and $f_s/f_b = 12$ whereas the value of b_{max} for $f_s/f_b = 16$ appears to be closer to the boundary value $4/\pi$.

In Table 2 the values of b^2 that are as close as possible to $4/\pi$ and Q_{cr} are shown for $n = 3, 4, 5$ and $f(k) = k^p$. Q_{cr} is the value for which,

$$\sum_{k=0}^{n-1} Q_{cr}^{k^p} = 2.$$

Other functions $f(k)$ could be considered that might give a more satisfactory b^2. However, the values of b^2 achieved for $f(k) = k^p$ with modest values of n are considered satisfactory for the purposes of this work.

It now remains to specify the parameter α. This constant concerns the optimization of S/N and is specified by the absolute signal levels to be applied to the modulator.

5 The Signal-to-Noise Ratio for the Various Cases of Interest

The signal-to-noise ratios are going to be examined for $n = 5$ and $n = 4$ and for three sampling rates of interest. For $f(k) = k^p$ equations (26) become:

$$\mu = \alpha\lambda \sqrt{\frac{2}{\pi}} \sum_{k=0}^{n-1} Q^{k^p} \zeta(k\tau)$$

$$m^2 = \alpha^2 \sum_{k=0}^{n-1} Q^{2k^p} + \alpha^2 \sum_{k=0}^{n-1} Q^{k^p} \sum_{\substack{l=0 \\ l \neq k}}^{n-1} Q^{l^p} c_{kl}.$$

1. $n = 5$

(a) $f_s/f_b = 8$

Taking the values of Q and p from Table 2 and $\zeta(k\tau)$, c_{kl}, λ^2 from Tables 1(a), (b) and (c), equation (27) becomes:

$$\left(\frac{S}{N}\right)_{(8)} = \frac{(x/\alpha)^2}{0.206(x/\alpha)^2 - 0.968(x/\alpha) + 2.28}. \quad (44)$$

Combining equations (44) and (17) we get

$$\left(\frac{S}{N}\right)_{(8)LP} = \frac{4(x/\alpha)^2}{0.206(x/\alpha)^2 - 0.968(x/\alpha) + 2.28}. \quad (45)$$

The plot of equation (45) is shown in Fig. 13 from which we can verify the wide dynamic range of this a.d.m. system.

Table 2

Values of b^2, Q_{cr} and DR for the three different sampling rates of interest and $n = 3, 4, 5$ ($f(k) = k^p$)

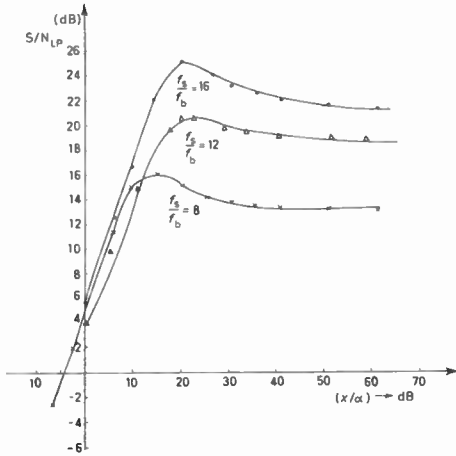| n	f_s/f_b	b^2	Q_{cr}	p	DR (dB)	$(S/N)_{LP}^{max}$ (dB)
3	8	1.11077	0.5	0	22.67	16.58
3	12	0.96693	0.5	0	15.78	22.83
3	16	0.93530	0.5	0	14.56	26.88
4	8	$4/\pi$	0.43	0.45	∞	15.90
4	12	1.1203	0.3333	0	23.27	21.82
4	16	0.99311	0.3333	0	16.83	26.37
5	8	$4/\pi$	0.43	0.64	∞	15.90
5	12	$4/\pi$	0.31	0.24	∞	21.18
5	16	1.15508	0.25	0	25.8	25.41

Fig. 13. Theoretical $(S/N)_{LP}$ vs. (x/α) for $f_s/f_b = 8$, $f_s/f_b = 12$ and $f_s/f_b = 16$ and for five-bit delay $(n = 5)$.

(b) $f_s/f_b = 12$

Working as before we get:

$$\left(\frac{S}{N}\right)_{(12)LP} = \frac{6(x/\alpha)^2}{0.0914(x/\alpha)^2 - 0.740698(x/\alpha) + 3}. \quad (46)$$

The plot of equation (46) is also shown in Fig. 13 where the wide dynamic range is confirmed.

(c) $f_s/f_b = 16$

The corresponding function for the signal-to-noise ratio is given by:

$$\left(\frac{S}{N}\right)_{(16)LP} = \frac{8(x/\alpha)^2}{0.05139(x/\alpha)^2 - 0.55(x/\alpha) + 2.66}. \quad (47)$$

The plot of this equation is also shown in Fig. 13.

2. $n = 4$

(a) $f_s/f_b = 8$

This case is of particular interest since the circuit fabricated in the laboratory employed this algorithm. We have:

$$\left(\frac{S}{N}\right)_{(8)LP} = \frac{4(x/\alpha)^2}{0.206(x/\alpha)^2 - 1.0278(x/\alpha) + 4.99}. \quad (48)$$

The graph of the above equation is shown in Fig. 14. From this graph we can see that the dynamic range is still wide.

(b) $f_s/f_b = 12$

Here, we have:

$$\left(\frac{S}{N}\right)_{(12)LP} = \frac{6(x/\alpha)^2}{0.0914(x/\alpha)^2 - 0.8(x/\alpha) + 3.08}. \quad (49)$$
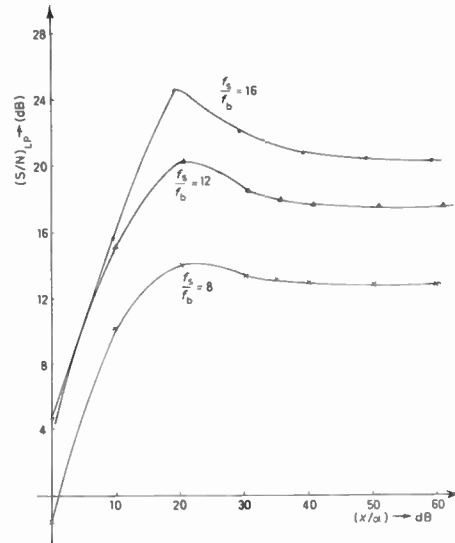
The plot of this expression is also shown in Fig. 14.

Fig. 14. Theoretical $(S/N)_{LP}$ vs. (x/α) for $f_s/f_b = 8$, $f_s/f_b = 12$ and $f_s/f_b = 16$ and for four-bit delay $(n = 4)$.

(c) $f_s/f_b = 16$

In this case we have:

$$\left(\frac{S}{N}\right)_{(16)LP} = \frac{8(x/\alpha)^2}{0.05139(x/\alpha)^2 - 0.65(x/\alpha) + 3.2}. \quad (50)$$

The plot of the above is also shown in Fig. 14.

6 Implementation of the Proposed A.D.M. System and Experimental Results

In this Section a brief description of the implementation of an a.d.m. system employing the optimum calculated algorithm in which $n = 4$ and $f_s/f_b = 8$ is given and the experimental results are compared with the theoretical results.

6.1 Implementation

Figure 15 shows a detailed block diagram of the circuit implementation. The input signal $x(t)$ is compared to the feedback signal $y(t)$; the output of the comparator (quantizer) is passed through a D-type flip-flop which acts as a sample-and-hold circuit. The two complementary outputs of the flip-flop are fed to two shift registers to form the binary signals c_{i-1}, c_{i-2}, c_{i-3} and $\bar{c}_{i-1}$, $\bar{c}_{i-2}$ and $\bar{c}_{i-3}$. These binary signals are then used to implement the algorithm described in the previous Section. According to this algorithm the step size at every instant should obey the following equation:

$$\Delta_i/\alpha = \sum_{k=0}^{3} c_{i-k} Q^{k^p}.$$

The values of Q and p are taken equal to 0.43 and 0.45 (see Table 2). The values that Δ_i/α takes on when $c_i = 1$
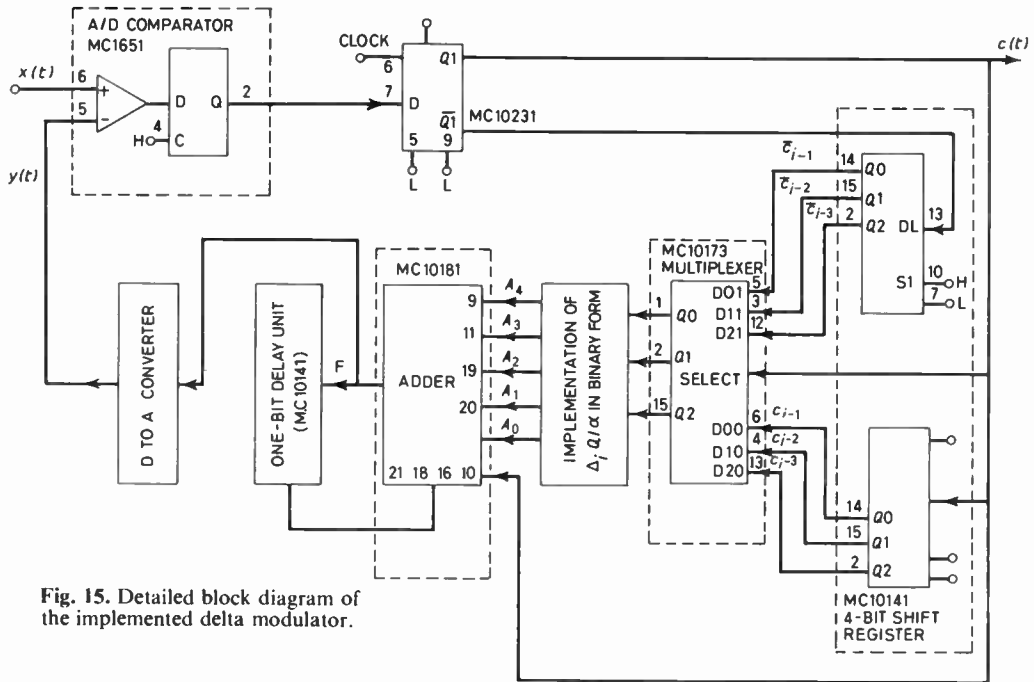
Fig. 15. Detailed block diagram of the implemented delta modulator.

are shown in Table 3 for the various states of c_{i-1}, c_{i-2}, c_{i-3}.

The parameter Δ_i^Q/α is the normalized quantized step with base 0.1 while A_4, A_3, A_2, A_1 and A_0 represent the values of Δ_i^Q/α in binary form.

The signals A_4, A_3, A_2, A_1 and A_0 (i.e. Δ_i^Q/α) pass through the binary adder (subtractor) causing the step to be added to (or subtracted from) the previous value of y. The output of the adder (subtractor) represent, in binary form, the tracking signal y. This binary signal is then converted to an analogue signal and fed to the inverting terminal of the comparator. It is noted that all the digital circuits used in the implementation of the a.d.m. system belong to the emitter coupled logic (ECL) family.
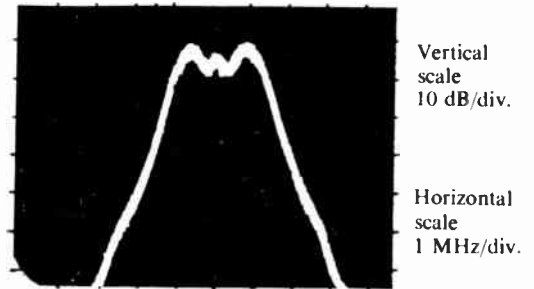
6.2 Experimental Results

The input to the coder was a Gaussian noise source with spectrum as shown in Fig. 16(a). Figure 16(b) shows the spectrum of the demodulated signal (after the low-pass filter of the decoder).

Table 3

Values of Δ_i/α for the various states of c_{i-1}, c_{i-2}, c_{i-3} and $c_i = 1$

c_{i-3}	c_{i-2}	c_{i-1}	Δ_i/α	Δ_i^Q/α	A_0	A_1	A_2	A_3	A_4
1	1	1	2	20×0.1	0	0	1	0	1
0	1	1	1.476	15×0.1	1	1	1	1	0
1	0	1	1.368	14×0.1	0	1	1	1	0
0	0	1	0.844	8×0.1	0	0	0	1	0
1	1	0	1.140	10×0.1	0	1	0	1	0
0	1	0	0.616	6×0.1	0	1	1	0	0
1	0	0	0.508	5×0.1	1	0	1	0	0
0	0	0	0	0×0.1	0	0	0	0	0

The bandwidth of the signal was 2 MHz ($f_b = 2$ MHz). The resultant output power was measured in selected 5 kHz bands with the aid of a wave analyser. This power comprises a representation of the input power with additional noise generated in the coder itself. A band-rejection filter was then inserted before the coder to block the applied signal in the frequency band where the measurement is made.

Vertical scale 10 dB/div.

Horizontal scale 1 MHz/div.

(a) Spectrum of the input signal.
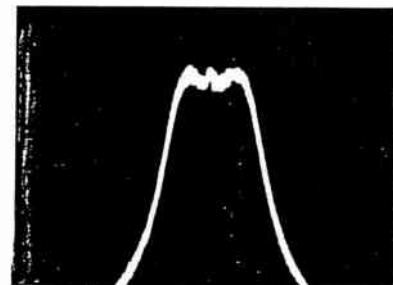
(b) Spectrum of the output signal.

Fig. 16. Experimental results for Gaussian noise signal.

The measured power is therefore the noise generated in the coder alone. A signal-to-noise ratio for the coder can be determined from these two measurements for the particular band of frequency chosen.

In Fig. 17 the experimentally measured $(S/N)_{LP}$ results are shown. These measurements were taken at 150 kHz.
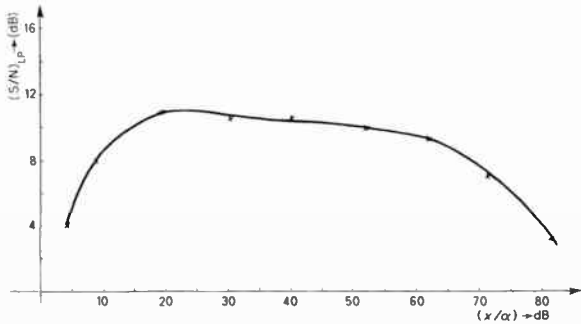

Fig. 17. Experimental $(S/N)_{LP}$ vs. (x/α) for $f_s/f_b = 8$.

Comparing the theoretical results with the experimental results (i.e. Fig. 14 $f_s/f_b = 8$ with Fig. 17) the following conclusions may be deduced:

1. The dynamic range is about 50 dB. This is not in agreement with the theoretical results where the dynamic range was predicted to be infinite. As will be illustrated, the occurrence of slope overload at high input signal levels limits the dynamic range.

2. The experimental values of the $(S/N)_{LP}$, for relatively small input signal levels differ from the theoretical by 2 dB. This is basically due to the fact that the tracking for the sampling rate used was not 'completely' optimum.

The dramatic reduction of the signal-to-noise ratio (and consequently of the dynamic range) which occurs for high values of (x/α) (where slope overload noise dominates rather than quantizing noise) is due to the fact that at high values of (x/α) the main assumptions (14) become invalid for the reason that a considerable number of 'ones' ('zeros') appears at the output of the system for decreasing (increasing) $x(t)$.

7 Conclusions

In this paper we have presented a general theory which seems to be valid for any optimum delta modulation system. Then, on the basis of the general theory we derived a useful expression for the signal-to-noise ratio for a specific adaptation logic and a normal density function input signal with flat spectrum. The theoretical results seem to be in agreement with the experimental results. Besides, the values for the signal-to-noise ratio obtained with this algorithm are very close to the values obtained by others who used more complicated algorithms.

8 References

1 De Jager, F., 'Delta modulation—a new method of p.c.m. transmission using the 1 unit code', *Philips Res. Rep.*, 7, pp. 442–6, December 1952.

2 Libois, L. M., 'Un nouveau procede de modulation codes la modulation en delta', *L'Onde Electronique*, 32, pp. 26–31, January 1952.

3 Van de Weg, N., 'Quantizing noise of a single integration delta modulation system with an N-digit code', *Philips Res. Rep.*, 8, pp. 367–85, October 1953.

4 Zetterberg, L. H., 'A comparison between delta modulation and p.c.m.', *Ericsson Technics*, 1, pp. 95–154, 1955.

5 Gersho, A., 'Stochastic stability of delta modulation', *Bell Syst. Tech. J.*, 51, no. 4, pp. 821–41, April 1972.

6 O'Neal, J. B., Jr., 'Delta modulation quantizing, noise analytical and computer simulation results for Gaussian and television signals', *Bell Syst. Tech. J.*, 45, no. 1, pp. 117–41, January 1966.

7 Jayant, N. S., 'Adaptive delta modulation with a one-bit memory', *Bell Syst. Tech. J.*, 49, no. 3, pp. 321–2, March 1970.

8 Basworth, R. H. and Candy, J. C., 'A companded one-bit coder for television transmission', *Bell Syst. Tech. J.*, 48, pp. 1459–79, July 1969.

9 Winkler, M. R., 'High information delta modulation', IEEE International Conf. Rec., Pt. 8, pp. 260–5, 1963.

10 Adate, J. E., 'Linear and adaptive delta modulation', *Proc. IEEE*, 55, no. 3, pp. 298–308, March 1967.

11 Goodmay, D. J., 'A digital approach to adaptive delta modulation', *Bell Syst. Tech. J.*, 50, no. 4, pp. 1421–7, April 1971.

12 Der Slepian, D., 'On delta modulation', *Bell Syst. Tech. J.*, 51, no. 10, pp. 2101–36, December 1972.

13 Cummiskey, P., 'Single-integration, adaptive delta modulation', *Bell Syst. Tech. J.*, 54, no. 8, pp. 1463–74, October 1975.

14 Zetterberg, L. H. and Uddenfeldt, H., 'Adaptive delta modulation with delayed decision', *IEEE Trans. on Communications*, COM-22, no. 9, pp. 1195–8, September 1974.

15 Papoulis, A., 'Probability, Random Variables and Stochastic Processes' (McGraw-Hill, New York, 1965).

16 Nitadori Kazuhkio, 'Statistical analysis of Δpcm', *Electronics and Communications in Japan*, 41, no. 48, pp. 17–26, February 1965.

9 Appendix 1: The relationship between the input process and its first-order difference

The input signal function is assumed to be normal and stationary with mean value not necessarily zero, a variance S_0 and a flat spectrum given by:

$$S(\omega) = \frac{S_0 \pi}{\omega_b}, \quad \text{for } \omega \leqslant \omega_b$$
$$S(\omega) = 0, \quad \text{for } \omega > \omega_b \tag{51}$$

where ω_b is the bandwidth of the input signal.

The autocorrelation of the input signal is given as the inverse Fourier transform of the power spectrum, or:

$$R(\tau') = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau'} d\omega$$

and in our case,

$$R(\tau') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_0 \pi}{\omega_b} e^{j\omega\tau'} d\omega = \frac{1}{2\pi} \int_{-\omega_b}^{\omega_b} \frac{S_0 \pi}{\omega_b} e^{j\omega\tau'} d\omega$$

or

$$R(\tau') = S_0 \frac{\sin \omega\tau'}{\omega\tau'}. \tag{52}$$

The autocorrelation of the derivative of $x(t)$, dx/dt is given by:

$$R_{x'x'}(\tau') = -\frac{d^2 R(\tau')}{d\tau'^2}. \tag{53}$$

Since $\xi(t) = \tau' \, dx/dt$, we may see that the auto-correlation of the first-order difference $\xi(t)$ is given by:

$$R_\xi(\tau') = E\left[\tau \frac{dx(t)}{dt} \cdot \tau \frac{dx(t+\tau')}{dt}\right] = \tau^2 R_{x'x'}(\tau'). \tag{54}$$

Equation (54) indicates that in order to find the autocorrelation of $\xi(t)$, it suffices to find $R_{x'x'}(\tau')$.

Differentiating equation (52) twice with respect to time, we obtain:

$$R_{x'x'}(\tau') = -\frac{d^2 R(\tau')}{d\tau'^2} = -\omega_b^2 \frac{d^2(\tau')}{d(\omega_b \tau)^2}$$

or

$$R_{x'x'}(\tau') = -\left[\frac{\omega_b^2 S_0}{(\omega_b \tau')^2}\right] \times$$
$$\times \left[2\frac{\sin \omega_b \tau'}{\omega_b \tau'} - 2 \cos \omega_b \tau' - \omega_b \tau' \sin \omega_b \tau'\right] \tag{55}$$

or

$$R_{x'x'}(\tau') = -\omega_b^2 S_0 \left[2\frac{\sin \phi}{\phi^3} - 2\frac{\cos \phi}{\phi^2} - \frac{\sin \phi}{\phi}\right] \tag{56}$$

where

$$\phi \equiv \omega_b \tau'. \tag{57}$$

Since

$$\left.\frac{\sin \phi}{\phi}\right|_{\phi=0} = \lim_{\phi \to 0} \frac{\sin \phi}{\phi} = 1$$

and

$$\left.\frac{\cos \phi}{\phi^2}\right|_{\phi=0} = \lim_{\phi \to 0} \frac{\cos \phi}{\phi^2} = -\lim_{\phi \to 0} \frac{\sin \phi}{2\phi} = -\frac{1}{2}$$

and

$$\left.\frac{\sin \phi}{\phi^3}\right|_{\phi=0} = \lim_{\phi \to 0} \frac{\cos \phi}{3\phi^2} = -\lim_{\phi \to 0} \frac{\sin \phi}{6\phi} = -\frac{1}{6}.$$

For $\tau' = 0$, equation (56) becomes

$$R_{x'x'}(0) = -\omega_b^2 S_0 [-2(1/6) - 2(-1/2) - 1] = \tfrac{1}{3}\omega_b^2 S_0. \tag{58}$$

Thus, the variance of the first-order difference is given by (see also equation (54)),

$$R_\xi(0) = \tau^2 R_{x'x'}(0) = \tfrac{1}{3}\tau^2 \omega_b^2 S_0$$

or

$$R_\xi(0) = \frac{4\pi^2}{3}\left(\frac{f_b}{f_s}\right)^2 S_0. \tag{59}$$

We have

$$E[\xi_i^2] = R_\xi(0)$$
$$E[x_i^2] = S_0$$

and therefore,

$$\frac{E[\xi_i^2]}{E[x_i^2]} = \frac{R_\xi(0)}{S_0} = \frac{4\pi^2}{3}\left(\frac{f_b}{f_s}\right)^2. \tag{60}$$

Introducing

$$\lambda^2 \equiv \frac{4\pi^2}{3}\left(\frac{f_b}{f_s}\right)^2. \tag{61}$$

Equation (60) becomes

$$\frac{R_\xi(0)}{S_0} = \frac{E[\xi_i^2]}{E[x_i^2]} = \lambda^2. \tag{62}$$

From equations (54) and (56), we obtain:

$$\zeta(\tau') = \frac{R_\xi(\tau')}{R_\xi(0)}$$
$$= 3[2 \cos \phi/\phi^2 + \sin \phi/\phi - 2 \sin \phi/\phi^3] \tag{63}$$

where $\zeta(\tau')$ is the correlation coefficient of $\xi(t)$ and $\phi \equiv \omega_b \tau'$.

10 Appendix 2: The correlation of the output signal $c(t)$

The input to the quantizer (comparator) is the signal $\xi(t)$. The mean value of $\xi(t)$ is given by:

$$E[\xi(t)] = \tau E[dx/dt] = \tau \frac{d}{dt}(E[x(t)]) = 0$$

because the mean value of $x(t)$ is constant.

The process $x(t)$ is normal and therefore the first-order derivative and the first-order difference are also normal. From equation (62), we have:

$$E[\xi^2(t)] = \lambda^2 S_0.$$

Thus, the probability density function of $\xi(t)$ is given by:

$$f(\xi) = \frac{1}{2\pi\sqrt{R_\xi(0)}}\exp\left(-\frac{\xi^2}{2R_\xi(0)}\right)$$
$$= \frac{1}{2\pi\lambda\sqrt{S_0}}\exp\left(-\frac{\xi^2}{2\lambda^2 S_0}\right). \tag{64}$$

The random variables

$$\xi(t+\tau') \quad \text{and} \quad \xi(t)$$

are jointly normal, with the same variance $R_\xi(0)$ and joint density given by:

$$f(\xi_1, \xi_2, \tau') = \frac{1}{2\sqrt{R_\xi^2(0) - R_\xi^2(\tau')}} \times$$
$$\times \exp\left[-\frac{R_\xi(0)\xi_1^2 - 2R_\xi(\tau')\xi_1\xi_2 + R_\xi(0)\xi_2^2}{2(R_\xi^2(0) - R_\xi^2(\tau'))}\right]. \tag{65}$$

The output of the quantizer will be:

$$c(t) = \begin{cases} 1 & \text{if } \xi(t) > 0 \\ -1 & \text{if } \xi(t) \leqslant 0. \end{cases}$$

The mean value of $c(t)$ is zero and its autocorrelation is given by:

$$E[c(t+\tau')c(t)] = 1 \cdot P[\xi(t+\tau')\xi(t) > 0] - $$
$$ - 1 \cdot P[\xi(t+\tau')\xi(t) \leqslant 0]. \tag{66}$$

In order to find the probabilities

$$P[\xi(t+\tau')\xi(t) > 0]$$

and

$$P[\xi(t+\tau')\xi(t) \leqslant 0]$$

we introduce the random variable

$$z = \xi(t+\tau')/\xi(t).$$

The region D of the $\xi(t+\tau')$, $\xi(t)$ plane such that,

$$\frac{\xi(t+\tau')}{\xi(t)} \leqslant z$$

is the shared area of Fig. 18, because

if $\xi(t) > 0$, then $\xi(t+\tau') \leqslant \xi(t)z$

if $\xi(t) \leqslant 0$, then $\xi(t+\tau') \geqslant \xi(t)z$.
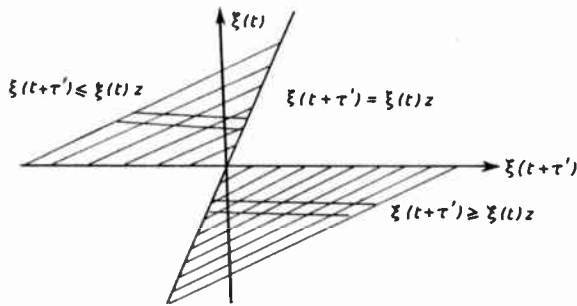

Fig. 18

The masses in D can be found by integration over the horizontal strips of Fig. 18.

$$F_z(z) = \int_0^\infty \int_{-\infty}^{\xi_2 z} f(\xi_1, \xi_2) \, d\xi_1 \, d\xi_2 + \int_\infty^0 \int_{\xi_2 z}^\infty f(\xi_1, \xi_2) \, d\xi_1 \, d\xi_2$$

where ξ_1 and ξ_2 are the values of $\xi(t+\tau')$ and $\xi(t)$ respectively.

Differentiating with respect to z, and employing the fact that

$$f(\xi_1, \xi_2) = f(-\xi_1, -\xi_2),$$

we obtain the following expression for $f_z(z)$.

$$f_z(z) = 2 \int_0^\infty \xi_2 f(\xi_2 z, \xi_2) \, d\xi_2. \tag{67}$$

Substituting into equation (67) the value of $f(\xi_2 z, \xi_2)$ as it is given by equation (65), we obtain:

$$f_z(z) = 2 \int_0^\infty \xi_2 \frac{1}{2\pi\sqrt{R_\xi^2(0) - R_\xi^2(\tau')}} \times$$

$$\times \exp\left[-\frac{R_\xi(0)\xi_2^2 z^2 - 2R_\xi(\tau')z\xi_2^2 + R_\xi(0)\xi_2^2}{2(R_\xi^2(0) - R_\xi^2(\tau))} \right] d\xi_2.$$

Noting that

$$\int_0^\infty e^{-\xi^2/2a^2} \, d\xi = a^2,$$

the expression for $f_z(z)$ becomes:

$$f_z(z) = \frac{\sqrt{1-\zeta^2}\, R_\xi^2(0)/\pi}{R_\xi^2(0)(z-\zeta)^2 + R_\xi^2(0)(1-\zeta^2)}$$

$$= \frac{\sqrt{1-\zeta^2}}{\pi[(z-\zeta)^2 + (1-\zeta^2)]} \tag{68}$$

where

$$\zeta \equiv \zeta(\tau') \equiv \frac{R_\xi(\tau')}{R_\xi(0)}.$$

Thus

$$P[\xi(t+\tau')\xi(t) > 0] = P[z > 0],$$
$$P[\xi(t+\tau')\xi(t) \leqslant 0] = P[z < 0]. \tag{69}$$

But

$$P[z \leqslant 0] = \int_{-\infty}^0 f_z(z) \, dz = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\zeta}{\sqrt{1-\zeta^2}}$$

or

$$P[z \leqslant 0] = \frac{1}{2} + \frac{1}{\pi} \sin^{-1} \zeta(\tau'). \tag{70}$$

This is so because

$$\tan^{-1} \frac{\zeta}{\sqrt{1-\zeta^2}} = \sin^{-1} \zeta.$$

For the $P[z > 0]$, we have

$$P[z > 0] = 1 - P[z \leqslant 0] = \frac{1}{2} - \frac{1}{\pi} \sin^{-1} \zeta. \tag{71}$$

On the basis of equations (69), (70) and (71), equation (66) becomes:

$$E[c(t+\tau')c(t)] = \frac{2}{\pi} \sin^{-1} \zeta(\tau'). \tag{72}$$

11 Appendix 3: The cross-correlation between $\xi(t)$ and $c(t)$

We are interested in specifying the cross-correlation between $\xi(t)$ and $c(t)$ or

$$R_{\xi c}(\tau') = E[\xi(t+\tau')c(t)]. \tag{73}$$

In order to find $R_{\xi c}(\tau')$, it is necessary to determine the joint distribution probability density between ξ and c, i.e. $f_{\xi c}(\xi, c)$. We have

$$P[\xi(t+\tau') \leqslant \xi, c(t) = -1] = P[\xi(t+\tau') \leqslant \xi, \xi(t) \leqslant 0]$$
$$= F(\xi, 0) \tag{74}$$

where

$$F(\xi, 0) = \int_{\xi_1 = -\infty}^{\xi} \int_{\xi_2 = -\infty}^{0} \frac{1}{2\pi\sqrt{R_\xi^2(0) - R_\xi^2(\tau')}} \times$$

$$\times \exp\left\{ -\frac{R_\xi(0)\xi_1^2 - 2R_\xi(\tau')\xi_1\xi_2 - R_\xi(0)\xi_2^2}{2[R_\xi^2(0) - R_\xi^2(\tau')]} \right\} d\xi_1 \, d\xi_2 \tag{75}$$

and

$$P[\xi(t+\tau') \leqslant \xi, c(t) = 1]$$
$$= P[\xi(t+\tau') \leqslant \xi, \xi(t) > 0]$$
$$= P[\xi(t+\tau') \leqslant \xi] - P[\xi(t+\tau') \leqslant \xi, \xi(t) \leqslant 0]$$
$$= F(\xi) - F(\xi, 0) \tag{76}$$

where

$$F(\xi) = \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi R_\xi(0)}} \exp\{-\xi^2/2R_\xi(0)\} \, d\xi. \tag{77}$$

By virtue of equations (74) and (76), equation (73) becomes:

$$R_{\xi c}(\tau') = E[\xi(t+\tau')c(t)]$$

$$= -1 \int_{-\infty}^{\infty} \xi \frac{\partial F(\xi, 0)}{\partial \xi} d\xi + 1 \int_{-\infty}^{\infty} \xi \frac{\partial F(\xi)}{\partial \xi} -$$

$$- 1 \int_{-\infty}^{\infty} \xi \frac{F(\xi, 0)}{\partial \xi}. \tag{78}$$

However,

$$\int_{-\infty}^{\infty} \xi \frac{\partial F(\xi)}{\partial \xi} d\xi = E[\xi(t)] = 0$$

and equation (78) becomes:

$$R_{\xi c}(\tau') = -2 \int_{-\infty}^{\infty} \xi \frac{\partial F(\xi, 0)}{\partial \xi} d\xi. \tag{79}$$

But,

$$\frac{\partial F(\xi, 0)}{\partial \xi} = \int_{-\infty}^{0} f(\xi, p) \, dp$$

where $f(\xi, p)$ is given by equation (75) for $\xi_1 \equiv \xi$ and $\xi_2 \equiv p$. Thus

$$R_{\xi c}(\tau') = -2 \int_{\xi=-\infty}^{\infty} \int_{p=-\infty}^{0} \xi \frac{1}{2\pi\sqrt{R_\xi^2(0) - R_\xi^2(\tau')}} \times$$

$$\times \exp\left\{ -\frac{R_\xi(0)\xi^2 - 2R_\xi(\tau')\xi p + R_\xi(0)p^2}{2[R_\xi(0)^2 - R_\xi(\tau')^2]} \right\} d\xi \, dp. \tag{80}$$

Denoting

$$A \equiv 2\pi\sqrt{R_\xi^2(0) - R_\xi^2(\tau')}$$

$$B \equiv 2R_\xi(0)\left[1 - \frac{R_\xi^2(\tau')}{R_\xi(0)} \right] = 2R_\xi(0)[1 - \zeta(\tau')^2].$$

Equation (80) becomes:

$$R_{\xi c}(\tau') = -\frac{2}{A} \int_{p=-\infty}^{0} \int_{\xi=-\infty}^{\infty} \xi \times$$

$$\times \exp\left[-\frac{\xi^2 - 2\zeta(\tau')\xi p + p^2}{B} \right] d\xi \, dp$$

$$= -\frac{2}{A} \int_{p=-\infty}^{0} \int_{\xi=-\infty}^{\infty} \xi \exp\left[-\frac{(\xi - \zeta(\tau')p)^2 + p^2(1 - \zeta^2(\tau'))}{B} \right] d\xi \, dp$$

$$= -\frac{2}{A} \int_{p=-\infty}^{0} \exp\left[-\frac{1 - \zeta^2(\tau')}{B} p^2 \right] \int_{-\infty}^{\infty} \xi \exp\left[-\frac{(\xi - \zeta(\tau')p)^2}{B} \right] d\xi \, dp. \tag{81}$$

But:

$$\int_{-\infty}^{\infty} \xi \exp\left[-\frac{(\xi - \zeta(\tau')p)^2}{B} \right] d\xi$$

$$= \int_{-\infty}^{\infty} (\xi - \zeta(\tau')p) \exp\left[-\frac{(\xi - \zeta(\tau')p)^2}{B} \right] d\xi +$$

$$+ \int_{-\infty}^{\infty} \zeta(\tau')p \exp\left[-\frac{(\xi - \zeta(\tau')p)^2}{B} \right] d\xi$$

$$= \int_{-\infty}^{\infty} \zeta(\tau')p \exp\left[-\frac{(\xi - \zeta(\tau')p)^2}{B} \right] d\xi$$

because

$$\int_{-\infty}^{\infty} (\xi - \zeta(\tau')p) \exp\left[-\frac{(\xi - \zeta(\tau')p)^2}{B} \right] d\xi = 0$$

since the function in the integral is odd. From a table of definite integrals, we have:

$$\int_{-\infty}^{\infty} \exp\left[-\frac{(\xi - \zeta(\tau)p)^2}{B} \right] d\xi = \sqrt{\pi B}.$$

Equation (81) therefore becomes:

$$R_{\xi c}(\tau') = -\frac{2\sqrt{\pi B}}{A} \int_{-\infty}^{0} \zeta(\tau')p \exp\left[-\frac{1 - \zeta^2(\tau')}{B} p^2 \right] dp$$

$$= \frac{2\sqrt{\pi B}}{A} \zeta(\tau') \int_{0}^{\infty} p \exp\left[-\frac{1 - \zeta^2(\tau')}{B} p^2 \right] dp$$

$$= \zeta(\tau') \frac{\sqrt{\pi B}}{A} \cdot \frac{B}{1 - \zeta^2(\tau')}.$$

Substituting the values of A and B into the above equation, we obtain:

$$R_{\xi c}(\tau') = \zeta(\tau') \sqrt{\frac{2R_\xi(0)}{\pi}}$$

and from equation (62),

$$R_{\xi c}(\tau') = \zeta(\tau')\lambda \sqrt{\frac{2}{\pi} S_0}. \tag{82}$$

From equations (72) and (82), we have:

$$c_{ik} = E[c_i c_k] = \frac{2}{\pi} \sin^{-1} \zeta(|i-k|\tau)$$

$$\tag{83}$$

$$R_{ik} = E[\xi_i c_i] = \zeta(|i-k|\tau)\lambda \sqrt{\frac{2}{\pi}} \sqrt{S_0}$$

where τ is the sampling period.

_ _

A transportable satellite ground station for television relay

B. SALKELD, C.Eng., M.I.E.R.E.*

D. GRIFFITHS, C.Eng., M.I.E.E.*

and

S. VERMA, B.Eng., C.Eng., M.I.E.R.E.*

SUMMARY

Since 1978 IBA has been experimenting with a
transportable satellite ground station designed to transmit
television signals via the OTS satellite using the 14 GHz
up-link frequency band. The aim of these experiments has
been to develop the use of transportable satellite up-links
for reporting news and events by satellite on a national
and international scale. The paper describes the
background to the requirement for such terminals and
reports on the technical results obtained.

The practical experience gained has made it possible to
define the desirable features of operational terminals of
this type and these are considered.

The future for television news reporting via satellite will
depend on the economics of this method in comparison
with others and on whether satisfactory technical and
regulatory methods can be devised to safeguard the use of
the geostationary orbit. The extent of these problems and
the possible solutions are considered.

* Independent Broadcasting Authority, Engineering Division,
Crawley Court, Winchester, Hants SO21 2QA.

1 The Requirement

Communications satellites offer the unique advantage of
being able to direct reliable high-quality signals to
virtually all locations within a large coverage area or
'footprint'. This represents the essential attraction of
satellites for television broadcasting to homes and for
'multipoint' distribution of information as in business
data systems. Conversely the ability to accept signals
from any location within the coverage can be exploited.

In broadcasting terms a suitable satellite system
represents a powerful means for contribution of news
material from the field to a base studio for compilation
into programme schedules.

Social changes in recent years have given rise to an
increasing proportion of 'foreign' news items in bulletins
and for long-distance transmission satellite methods can
be indispensable. Even for short distances the sheer
convenience of the satellite method can offset fairly high
costs.

It is clearly impossible to anticipate every source
location of programme material and even from locations
regularly used for sport or current affairs it is generally
uneconomic to provide permanent transmission capacity
to cope with the occasional peaks of demand.
Temporary additional capacity is often desirable and
demountable stations have occasionally been used.[1] A
satellite 'link-up' transmission terminal sufficiently small
and flexible to be easily moved on to location and put
into operation reasonably quickly could establish such
temporary connections with the minimum of forward
planning and virtually independent of geographic
distances.

With this background the Independent Broadcasting
Authority constructed a transportable ground transmit
station (Fig. 1) for experiments with the OTS satellite in
order to determine the viability of the scheme for
television news gathering. The relatively high OTS
frequency band permitted a much smaller design than
had hitherto been possible and the WARC '79
conference would offer an opportunity for 'mobile'
frequency allocations to be sought. The terminal was
first demonstrated at the International Broadcasting
Convention at Wembley in September, 1978 and since
then has been used for a series of tests and pre-
operational trials in a variety of circumstances.

2 The OTS Satellite

The experiments were made possible by the availability
of the OTS satellite. OTS, the Orbital Test Satellite, built
by the European Space Agency was launched in May
1978 for experimental purposes. Operation of OTS is the
responsibility of the Eutelsat organization and IBA
participation was arranged through the co-operation of
the British Post Office. OTS uses the 14 and 11 GHz
communication frequency bands for the up and down
links respectively. It carries two orthogonally polarized
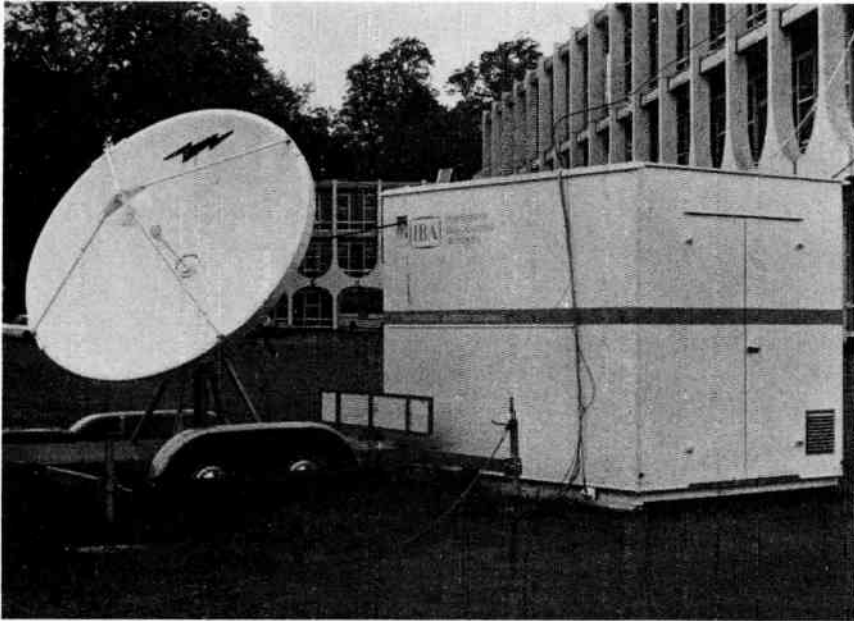
B. SALKELD, D. GRIFFITHS and S. VERMA

Fig. 1. The IBA transportable up-link

transponder channels of 120 MHz bandwidth coupled to a 'spotbeam' aerial and two channels of 40 MHz bandwidth driving a 'Eurobeam' aerial. Figure 2 shows the normal coverage footprint of each aerial. The spotbeam aerial gain is 35 dBi at the beam centre and the Eurobeam gain is 26 dBi.

All up-link transmission is via a Eurobeam receive aerial and satellite transponders with effective noise figure of 5 dB use travelling-wave amplifiers to produce a maximum saturated output of 20 watts for each channel. Transponder gain adjustment with a total range of 12 dB is available.
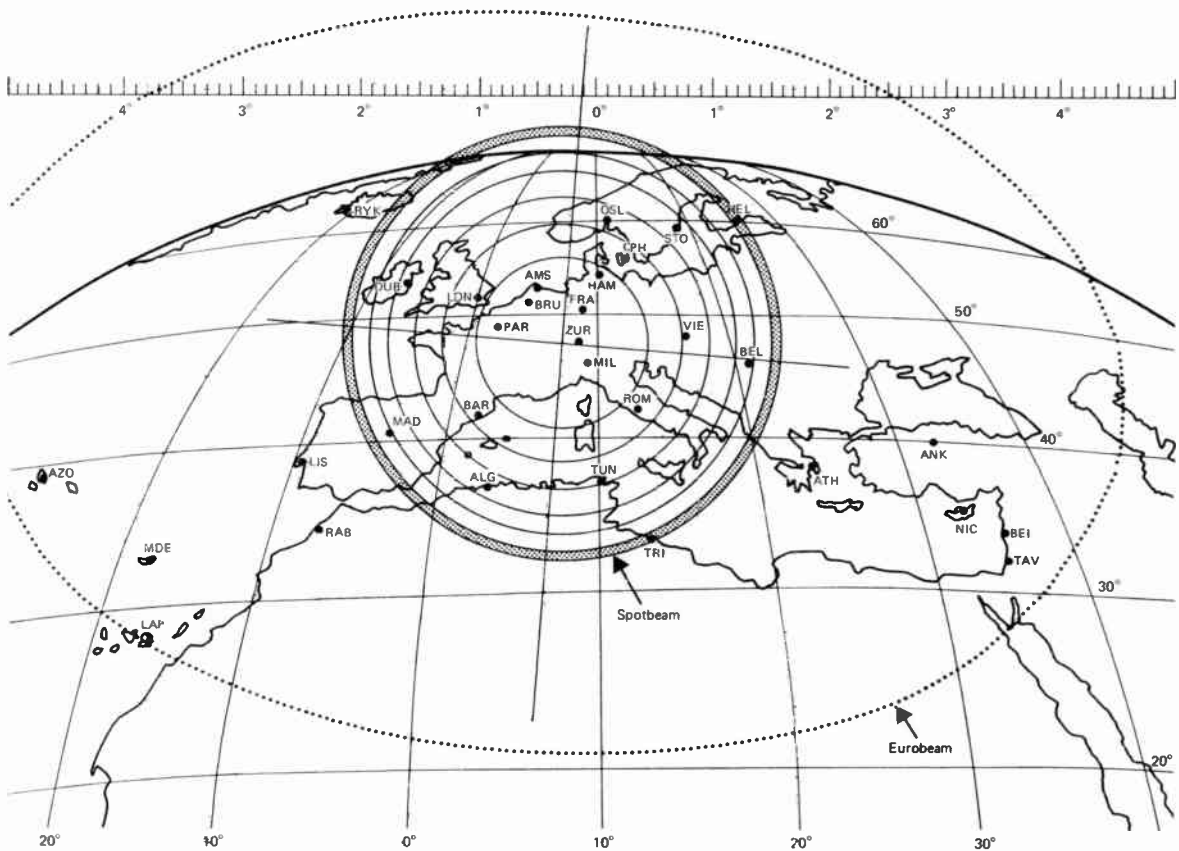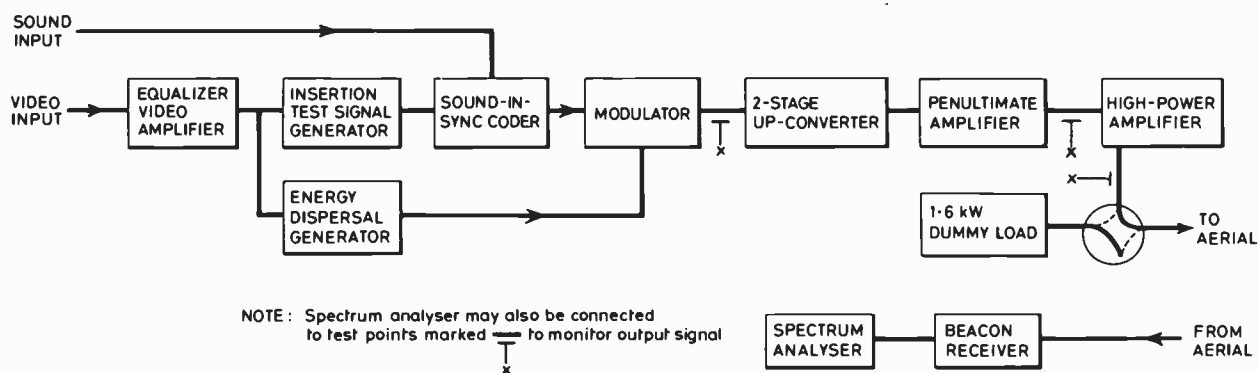

Fig. 2. OTS coverage

Fig. 3. Block diagram of up-link equipment

OTS is the forerunner of the European Communication Satellite system, ECS, which is intended to supplement the terrestrial networks in Europe and to provide additional flexibility for the Eurovision network of EBU.

3 The Equipment

The experimental up-link terminal consists of a small cabin containing transmitting equipment and a 2·5 metre diameter aerial mounted on a trailer. The complete assembly can be transported by a single vehicle which carries the equipment cabin and tows the trailer mounted aerial. The vehicle is equipped with a special mechanism enabling the cabin to be collected and deposited on site without recourse to cranes. The vehicle can then be released for its normal duties during the period of operation.

The equipment cabin which measures approximately 3 m by $2\frac{1}{4}$ m and $2\frac{1}{2}$ m high contains a standard Earth station high-power amplifier with $1\frac{1}{2}$ kW output and a single rack containing the modulator, up-converter and auxiliary equipment. A simplified block diagram is shown in Fig. 3 and the functions are described below.

The aerial reflector is stowed in a vertical position during transportation. Once at the chosen location the feed and sub-reflector are fitted and the trailer is stabilized by four extending arms with simple screw-down jacks.

3.1 Baseband Equipment

Vision and sound signals are accepted by the terminal on separate input sockets. The incoming video cable terminates in an equalizing distribution amplifier with a differential input to minimize the effects of hum and interference. The output from this equalizer is routed through an insertion test signal generator which adds test waveforms on lines 17, 18, and 330, 331, the accepted practice for international exchanges of programmes.

The sound programme is fed direct into a 'sound in sync' (s.i.s.) encoder which generates pulses during each line synchronizing period of the vision signal. The

amplitude of the s.i.s. pulses is limited to 0·7 V in order to contain the r.f. bandwidth of the subsequent frequency modulated spectrum.

In order to reduce the interference created by the spectral energy spikes inherent in all f.m. television signals, an energy dispersal tone is added to the composite video signal from the s.i.s. coder. A 25 Hz triangular waveform is used which is normally locked to the video picture frequency, but in the absence of a video signal the dispersal tone will free run so that the energy in the unmodulated carrier is spread over a discrete band.

3.2 Modulation

Frequency modulation takes place at 70 MHz with standard pre-emphasis (CCIR Recommendation 405–1). A 25 MHz peak-peak deviation is normally used occupying a nominal 40 MHz i.f. bandwidth.

3.3 Up-conversion

Up-conversion from the first i.f. of 70 MHz to the transmit frequency at 14 GHz takes place in two stages, with the second i.f. fixed at 750 MHz. The second conversion to 14 GHz uses a coaxial mixer followed by a waveguide bandpass filter. The local oscillator is a phase-locked source with two external reference oscillators corresponding to the band centre in the OTS satellite transponders. To change between these two channels it is necessary to change the up-converter output waveguide filter. Transmission on other frequencies within either transponder channel can be achieved using an external reference for the phase-locked source. The output power from the up-converter is approximately 1 mW.

3.4 Penultimate Amplifier

The output high power amplifier requires 100–200 mW of drive for full output. For reasons of convenience and availability this is provided by a 10 W travelling wave amplifier fitted with waveguide attenuators. The final transmit power can be set by varying the input attenuator. It is envisaged that a solid-state amplifier would be used in future.

3.5 High Power Amplifier

The output amplifier uses a five-cavity klystron capable of producing 1·5 kW at the cabinet output. Forced-air cooling is used with ducts directly to the outside through the end wall of the cabin. This makes for a simple and compact working arrangement suitable to the mobile application which would be difficult to achieve with an equivalent water-cooled device.

The klystron amplifier has an instantaneous 1 dB bandwidth of 80 MHz. The frequency of operation can be changed easily by selecting one of six pre-set tuning positions so that operation throughout the 14·0–14·5 GHz band is possible. This arrangement was preferred to a broadband travelling-wave tube for reasons of simplicity, particularly in the power supply. A series of interlocks protect the tube against flashover in the output waveguide due to dust or moisture.

The mains supply requirement is 3-phase, 415 V nominal with a maximum demand of 13 kVA. The amplifier is mounted in the cabin on six heavy-duty shock mountings.

The output from the amplifier is fed to a waveguide switch so that the output can be routed to the aerial or a dummy load. The dummy load is convection cooled and consists of four separate 400 W attenuator sections.

3.6 Beacon Receiver

The terminal contains a simple receiver for the satellite's beacon signals to enable alignment of the aerial prior to transmission. Two beacon signals are transmitted: BO with circular polarization and maximum e.i.r.p of 32 dBW, and TM linearly-polarized with maximum e.i.r.p. of 19 dBW. The output from the receive aerial feed is connected directly to a coaxial mixer. The mixer is fed from a fixed frequency local oscillator and produces a broad-band intermediate frequency without image rejection. The noise factor of this converter is about 10 dB. An i.f. in the range 0–500 MHz is connected to a

spectrum analyser which is used to display the beacon signals. To enable the analyser to be used over a narrow band when searching for beacon signals, the receiver generates its own markers close to and at a similar level to these received from the satellite (see Fig. 4). For OTS reception a single marker oscillator is used, arranged so that the fundamental of the marker at 70·3 MHz comes close to one beacon and the third harmonic of the marker oscillator at 211 MHz is close to the other; the beacons are spaced 211 MHz apart.

3.7 Aerial

To be reasonably transportable the aerial diameter must be no greater than that needed for adequate e.i.r.p., whilst for a given e.i.r.p. the off-beam interference will increase in direct proportion to the reduction in gain as a result of the higher transmitter power needed. For road transport a 2·5 metre diameter dish is a convenient compromise and the aerial chosen is a paraboloid reflector fitted with a Cassegrain feed and sub-reflector. The sub-reflector is mounted from the edge of the reflector by four support arms.

The aerial operates with linear polarization, crossed between transmit and receive. The feed consists of an orthogonal mode transducer (o.m.t.) to combine the transmit and receive ports, and a conical waveguide horn. The aerial complete with simple lead screw elevation adjustment rotates in azimuth on a central column fitted with a friction dish and clamp arrangement. In operation the trailer is stabilized by four extending arms with screw-down jacks at the end. This arrangement also helps to level the trailer platform under abnormal terrain conditions. Operation up to 130 km/h (80 m.p.h.) windspeed is achieved and no form of satellite tracking is needed since the positional stability of the satellite keeps it within 0·6° beamwidth of the up-link aerial.

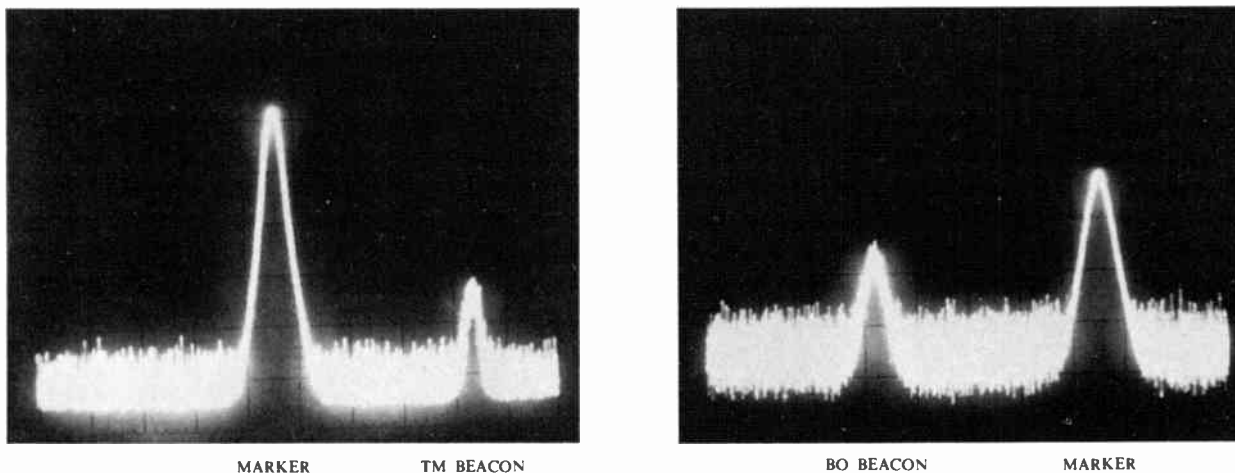MARKER   TM BEACON   BO BEACON   MARKER

Fig. 4. Beacon reception. Spectrum analyser display showing the satellite beacon signals appearing alongside marker pulses for identification. Aerial alignment is achieved by optimizing the circular and linear polarized beacon signals in turn.

4   Setting Up

On arrival at the operating site the antenna is pointed in approximately the right elevation and azimuth using angles calculated from the latitude and longitude of the site. More accurate setting follows using the satellite beacon signals.

To set the aerial initially an inclinometer, attached to the aerial reflector, is used for elevation and a compass bearing, normal to the direction of shoot, is taken across the front face of the dish for azimuth. A spectrum analyser connected to the output from the beacon receiver is placed on the trailer platform within view of the operator. The aerial is then panned, first in azimuth, then in elevation. A small azimuthal adjustment is normally sufficient to bring the satellite beacon signal within view on the analyser.

When operating with OTS the circularly polarized beacon BO is received first as this is the strongest signal and requires no rotation adjustment of the linear feed. Having acquired the satellite signal the linearly polarized telemetry beacon TM is measured and the feed rotated precisely to the wanted polarization so as to avoid interference with transmission on the opposite satellite channel. The aerial is then locked into position ready for transmission. Where the equipment cabin and aerial are located close together, connection between transmitter output and aerial feed is arranged by means of rigid lengths of waveguide jointed together by short lengths of twistable flexible guide. For roof mounting the aerial may be separated from the cabin by up to about 15 metres in which case a length of semi-rigid elliptical waveguide is used.

Off-loading and arranging the mains supply, establishing communications and setting up the aerial normally take about an hour, following which transmission can take place after the 5 minute warm-up time of the high-power amplifier.

In order to obtain optimum noise performance over the satellite link the satellite is normally set to maximum gain. This also has the effect of reducing the power needed from the up-link and the possibility of interference to other satellites. This is an important feature of satellite design when occasional small terminal operation is intended.

It is imperative not to drive the satellite beyond saturation but operation well below the input level needed for saturation is possible without any discernible loss of picture quality. This is partly because of the highly non-linear transfer characteristics of the satellite output amplifier, and partly because of the wide deviation used which results in high-quality pictures for all carrier levels above threshold.

5   Results

The performance of the terminal has been proved by a series of tests of the individual items and by system tests through the satellite under a range of operating and climatic conditions. System tests included sweep frequency gain and delay measurements using a conventional microwave link analyser and measurements of the vision and sound performance with a range of transmit power through the satellite. Both 625-line and 525-line transmissions took place and for the sound channel sub-carriers at 7·5 MHz or 6MHz were used as well as sound in sync. Operation took place from city centre locations, from open country and from an off-shore oil platform in the North Sea. It was regarded as equally important to obtain measured results of the performance and to assess the practical problems in an operational environment.

Reception usually took place using either the British Post Office 19 m diameter terminal at Goonhilly with G/T performance of 40 dB or at the IBA Winchester Engineering Centre using a receive-only installation with a 3 metre aerial (G/T 25 dB). Tests also took place using various European major ground stations operating with OTS, notably Fucino in Italy and on several occasions using a transportable 3 metre receive terminal owned by Ferranti.

Transmission took place in each channel of the OTS satellite using the spotbeam and Eurobeam satellite aerials. For some tests transmission of television signals in both directions using a single transponder took place with one signal sent from Goonhilly and the other from the IBA terminal. Simultaneous transmission of co-channel signals through opposite polarization transponders was frequently necessary in order to share and make best use of the available satellite transmission time.

Automatic measurements were taken continuously during some periods of transmission, but no significant variation in performance was observed except for variation in noise level with propagation conditions. Deliberate reduction in up-link transmit power affected only the signal-to-noise performance until threshold was reached.

The performance met the CCIR requirements for long distance terrestrial or satellite transmission (CCIR Recommendation 567) and the results were comparable with a single inter-city link even though the overall geographic span could represent many such links in tandem. Subjectively, except when the up-link power was deliberately reduced, the tests resulted in picture and sound quality indistinguishable from the source material. This was especially striking when sent and received signals were displayed side by side and the time delay (about 0·3 second) via the satellite could be seen.

The most informative numerical results are present in Table 1 and Figs. 5 and 6. These measurements were taken automatically for prolonged periods whilst an insertion test signal was being radiated. No variations were noted, except for noise, in relation to propagation conditions.

5.1 Overall Vision and Sound Tests

The performance obtainable with the terminal is determined almost entirely by the sensitivity of the ground receiving system and that of the satellite itself. Figure 5 shows the measured noise performance with various combinations of satellite and ground receive station aerial.
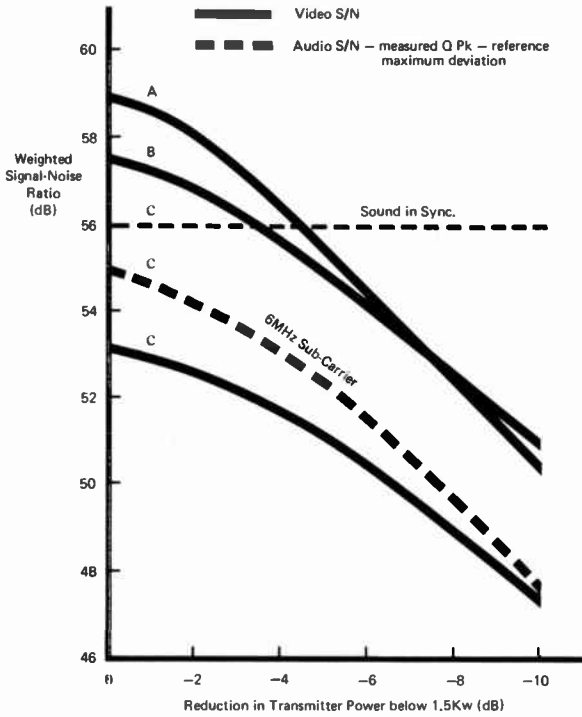

Fig. 5. Noise performance

Satellite Transmit Aerial	Earth Receive Aerial	OTS Channel
A Eurobeam	19 m diameter	2
B Spotbeam	19 m diameter	4
C Spotbeam	3 m diameter	4

It can be seen that, in the UK with the OTS satellite, up-link transmit power levels as low as 150 W gave very acceptable picture quality from a major 19 m receive terminal. This would confirm that operation from the extreme edge of the satellite Eurobeam coverage zone is possible with an up-link of the existing 1·5 kW design. For operation in central Europe a power level of 500 W would certainly be practical, and allow an operating margin for rainfall.

Table 1

Vision performance of the mobile up-link operating via OTS into 3 metre diameter receive terminal

Video Parameter	Satellite Channel 4
luminance non-linearity (%)	1·5
differential gain (%)	3·2
differential phase (degrees)	1·0
chrominance – luminance crosstalk (%)	0·6
chrominance – luminance gain inequality (dB)	+0·7
chrominance – luminance delay inequality (ns)	+11
2T pulse K rating (%K)	1·0

Considering the receive terminal, it can be seen that there is not a great change in performance between a 3 metre and a 19 metre terminal in average conditions. In heavy rainfall conditions the signal from a small terminal would fall below the f.m. threshold but the proportion of time for which this occurs is small and for occasional use such as news gathering purposes a
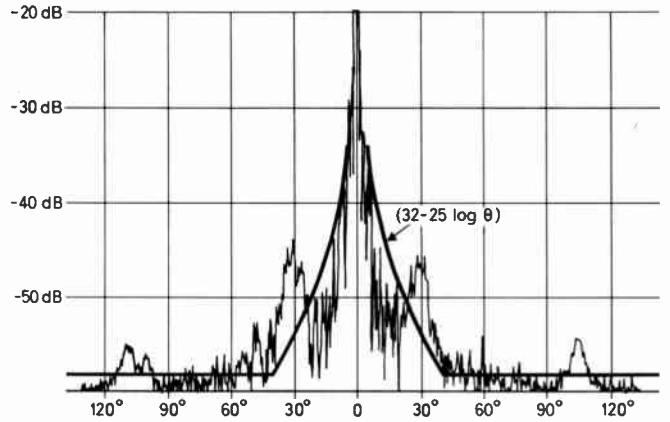

Fig. 6. Measured aerial performance

terminal of, say 5 metres diameter could be located at studio premises. Figure 7 shows the estimated availability for operation in the UK with a 2·5 metre, 500 W transmit terminal and a 5 metre diameter terminal, 30 dB/K receive terminal and a satellite equivalent to OTS. This type of operation appears to have economic advantages and the elimination of a long terrestrial back-haul link from a distant ground station would improve the reliability of the whole connection.

Apart from noise, the video performance (Table 1) obtained was largely governed by the amplitude and phase linearity of the i.f. and r.f. sections of the link. For the experimental terminal, the modulator was a conventional microwave link design, over-driven to produce wide deviation and only first-order correction was made of its derivative linearity and group delay. The modulator performance accounted for most of the differential phase and gain distortion of the link. No attempt was made to correct the overall group delay of the path through the satellite and this had a secondary effect on the performance noticeable when using sub-carrier techniques for the sound channel. It was particularly notable that no significant impairments were attributable to driving the up-link output klystron into the non-linear portion of its characteristic.

5.2 Aerial Performance

The aerial was tested on a range to check its gain and radiation pattern in the wanted and unwanted polarization. Figure 6 shows a typical radiation pattern. The gain is approximately 48·5 dB, representing an efficiency of 53% and would be improved with a closer profile accuracy. The polarization discrimination

achieved using the simple beacon receiver for optimization has been measured at −30 dB which is adequate for frequency re-use. The radiation pattern is typical of many practical aerials at present in use
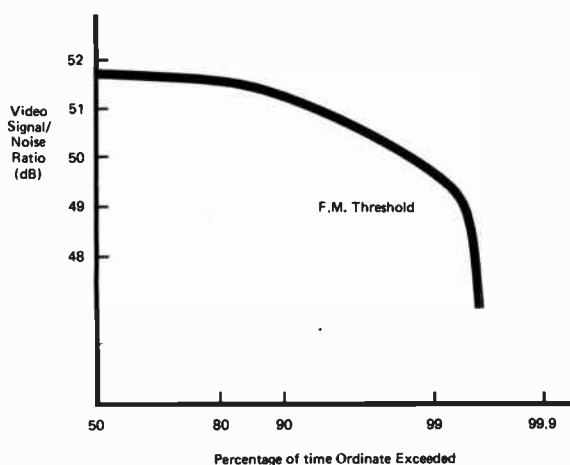

Fig. 7. System availability with small aerials.
Calculated performance for UK of a satellite 'news' link using a 500 W 'mobile up-link' with 2·5 m diameter aerial with reception on a 5 m diameter terminal of 30 dB/K performance.
Satellite of OTS characteristics is assumed.

although it does not meet the current CCIR Recommendation, (Recommendation 465-1) which was not in existence when the designs were introduced. Subsequent designs for operational terminals would perhaps need to be to a more stringent specification and an offset design may be beneficial if it can be realized at a favourable price.

5.3 Radiation Intensity
In order to operate a high-power radio transmitter in a useful variety of circumstances the possibility of unsafe radiation levels must be safeguarded. Extensive measurements of the radiation intensity around the IBA terminal showed that no level over 10 mW/cm² occurred except in the area directly in front of the aerial. This is the level at present considered safe for continuous exposure.

6 Applications
The experimental terminal has been operated from the IBA Engineering Centre at Winchester, from a number of demonstration sites in the UK and Europe, from events in Jersey and Ireland and from an oil rig in the North Sea. Several other applications have been proposed and investigated in some detail but for one reason or another the events have not taken place. These locations varied from the Sinai desert to the Outer Hebrides in Scotland and included possible transmissions

from Teheran and from Moscow for the Olympic Games.

In all of these actual or planned applications the use of the terminal has been found to be a practical engineering solution to the demands of the situation. The only requirements arising which could not be met were the use on board moving vessels at sea which would require a satellite tracking facility that has not been incorporated.

In taking advantage of the terminal for these events it was necessary to have sufficient advance notice to carry out the international frequency clearance procedure, a process normally taking several months. For the experimental period, due to the co-operation of the Administrations involved this was not a serious limitation for events planned a few weeks in advance, but it would be necessary in future to devise arrangements for prior clearance to be obtained so that the available channels may be known in advance. This would enable transportable terminals to be used at short notice for news events, which was not possible at all during the experimental period.

Apart from the possibility for live coverage of news stories from location, which is seen as the primary application for operational terminals, the experimental uses were found to arise in the following circumstances.

6.1 Locations from which no other connection is possible
Island sites may be too distant from the mainland for terrestrial television links to be feasible. The Channel Islands are in this situation being served by submarine cables and narrowband radio links not suitable for television transmission. This applies to a number of other British islands. Off-shore oil rigs, such as those now operating in the North Sea, are also often beyond reach of normal television microwave links.

6.2 Locations from which terrestrial links may be set up, but which are easier to deal with by satellite
This applies to many situations but in particular where several temporary microwave hops would be involved. The intermediate repeaters must have power supplies, they require aligning and may possibly need to be staffed during operations. A survey is often needed before the feasibility of the connection can be determined and if the terrain is difficult, the necessary sites may not be accessible by normal transport. Planning and implementing such links becomes a fairly large exercise involving a dozen or more people. The experimental terminal has been operated by a team of two plus a driver for the special vehicle and feasibility depends simply on finding a location with line-of-sight to the satellite. None of the sites investigated presented any difficulty in this respect even where steep mountain peaks surrounded the site as occurred in the proposed link with Sinai and in Switzerland.

6.3 Locations from which links exist but not in sufficient quantity to meet a peak demand

This situation can arise with almost any news or sporting event of significance. The visit of His Holiness the Pope to Ireland gave rise to such a problem even though high capacity television links and routes are available. The Moscow Olympic Games and similar events always give rise to severe pressure on the permanent communication links.

Television broadcasters generally seek to present their productions in a particular style and to do this effectively demands access to a link, separate from the 'multilateral' output normally made available by the host broadcaster. When an event has wide interest the demands for the available circuits can occupy a great deal of planning and negotiation for some time before the event.This may take a year for an event on the scale of the Olympic Games. An additional satellite circuit gives access to many reception points in the coverage zone and can be time-shared to meet a high proportion of these transmission needs.

6.4 'Invisible' Traffic

It was noted in several applications where the satellite link was made available for one purpose that the attraction of a straightforward high-quality link encouraged additional use for special transmissions not foreseen prior to the event, and in addition to the purpose for which the arrangements had been made. This appears to be an example of a well-known phenomenon where the provision of a communication channel creates its own demand and this can be helpful in defraying costs.

7 Operational Designs

Arising from the results of the technical test programe and the outcome of the various trials, a number of practical points were noted relevant to future designs for operational mobile up-links.

Some of these suggestions could conflict and the overriding needs are for a terminal able to get to site easily with good communication to base.

7.1 Transportation

For air transport the height of the experimental terminal was a limitation. Considerable scope existed to reduce the size and standard aircraft 'igloo' containers are available which could form a convenient basis for design. For shorter distances a single self-contained vehicle appears to be practical.

7.2 Radiated Power

The experimental terminal produced sufficient power to operate from the extreme edge of the satellite's Eurobeam coverage area. Significant savings in size and weight appear to be possible with a small reduction in r.f. output power and coverage zone. For operation within central Europe, a power level of a few hundred watts would be adequate. The mains power supply could then be single phase and perhaps self-contained. The requirement for three-phase power was not found to be a limitation during the trials described.

7.3 Communications

Good communications with the mobile terminal are vital to a smooth operation. For all the experimental work public telephone connection was used and was adequate for the circumstances encountered. Contact is necessary with the receiving ground station, with the base studio and with the satellite control centre. Both-way speech channels via the satellite appear to be the best way to achieve this, especially if the reception terminal were sited at the studio. Separate h.f. radio channels would be convenient if channels were available.

7.4 Input Conditions

Mains hum and interference on the temporary video cables feeding the up-link from camera or recorded sources in the environment of a busy news gathering activity can seriously affect the quality of the transmitted pictures. This was a frequent difficulty during the experimental phase, awkward to deal with systematically and wasteful of time on site. The effects may be visible only at receive points many miles distant, and communications may be poor. A simple analogue optical fibre connection to the up-link would be free from this sort of difficulty.

7.5 Reliability

The reliability of the equipment was found to be high despite the hazards of transportation and periods standing idle in the open without heating or dehydration. No redundancy was employed. Initially high tension leads in the power amplifier repeatedly fractured after the vibration of transport, but a replacement design has not failed.

The few faults encountered were all the result of mechanical stress rather than electrical failure. Elementary shock mounts were used and further attention to this aspect would appear to be worthwhile.

7.6 'Reverse' Vision

In order to monitor the up-link performance a simple receiver to display the satellite output signal would be achievable at modest additional cost. An elementary check of transmission quality could then be made and this would help remove ambiguity about the nature of any impairments reported from the home studio. It would also be possible to check when the satellite was clear for transmission and whether saturation was occurring.

7.7 Multiple Sound Channels

For international transmission a single video channel with a number of additional sound channels has application. Sound in sync together with an f.m. sub-carrier sound channel may often be sufficient, but multiple sound channels using a digitally modulated sub-carrier form an alternative.

8 Frequency Clearance

The international frequency table was amended by the WARC '79 decisions to permit land-mobile to satellite systems on a secondary basis in the 14–14·5 GHz frequency band, sharing with the primary space and terrestrial services. The way in which this will be administered is under consideration. In order to cut out the delay of an international clearance procedure for fixed terminals, one approach could be to determine in advance for each geographic zone the frequencies not used by terrestrial links and for subsequent operation to be permitted provided that actual interference was not caused. In determining the practical range of interference due account could be taken of the occasional nature of transmissions. This would enable a reasonable choice of channel frequencies for the 'hot news' situation whilst not restricting the future introduction of new permanent services.

The possibility of interference to adjacent satellites would also have to be safeguarded. Smaller transmit aerials carry the risk of greater off-beam interference as the r.f. power is increased to offset reduced aerial gain. Whilst the satellite orbit is relatively uncrowded at these frequency bands, channel frequencies to be used for small terminals may be chosen not to coincide with the interference-sensitive channels of nearby satellites. This process would be eased further if satellites of greatly different sensitivity were separated in orbit position.

The characteristics of occasional use ease the problem. For permanent links the economic considerations of the system design reliability determine the up-link power which is used. The satellite itself is the greatest cost and the best use is made of its available power by ensuring that the noise contribution from the up-link is negligible. High up-link power is normally available. The down link can then be engineered to achieve a continuity of service governed largely by down-link fading.

For occasional traffic the economics are quite different. Spare transponders or even whole satellites necessarily exist for redundancy and can be put to practical use without the constraints necessary for full time traffic. Rather lower up-link power is needed and if the satellite gain is increased during these transmissions the corresponding increase in incoming interference may well be an acceptable risk for this type of transmission.

It is not essential to saturate the satellite and the signal need be no more that that necessary to operate above the f.m. threshold at the receive station. The f.m. improvement effect of wide deviation is then sufficient to result in good quality vision and sound.

Further reduction in the interference risk can be achieved by energy spreading. In full-time use the amount of triangular dispersal signal which can be employed is usually determined by the need to remain within the defined channel bandwidth. When a whole wider-band transponder is available for occasional use and the television signal is the sole traffic, a larger dispersal signal can be employed as necessary.

Eventually, if these measures may become insufficient for a closely-packed satellite orbit it may be necessary to introduce digital modulation for the up-link. This offers a powerful tool to offset the energy spikes normally present in the f.m. television energy spectrum, and would represent an effective form of energy dispersal.[2]

9 The Costs

The probable costs of an operational service using transportable terminals depend on many factors which are difficult to quantify, but some guidance may be obtained from the following.

The capital cost of the up-link terminal in 1978 was under £100,000, the high-power amplifier forming the most expensive single item. It could be operated by a team of two plus a driver.

The cost of a 3 metre diameter 25 dB/K terminal for reception was about £30,000 at that time.

Full-time satellite transponder rentals are at present in the order of £1M per annum, equivalent approximately to £120 for each hour. Rental may be paid on a full-time basis and shared between users as in the Eurovision system. Spare time is available for occasional traffic and the overall use of the transponder becomes more efficient.

The rates for occasional transponder hire vary depending upon competitive forces and the policies of the operating agencies. The rental charges can be set to attract the optimum proportion of occasional use.

10 Conclusions

A small transportable satellite ground station for television news reporting has been described and the results of trials have been reported. A number of practical features for future designs have been noted and possible arrangements for frequency clearance have been indicated.

The notable features of the trial can be summarized as:

(1) The television performance of the up-link terminal with OTS met the CCIR specification for a hypothetical reference circuit and the targets for satellite transmission (Recommendation 567). This was achieved both with reception at a major 19 metre diameter station and at a 3 metre diameter station under average propagation conditions.

(2) Transportation by road or sea to site was found to be straightforward.

(3) The terminal could be aligned easily following arrival at site and operations started within about an hour. Reliable performance was then possible for several days at a time without further attention.

(4) 625-line and 525-line standard colour television signals were equally usable from city centres, open country or on an oil platform at sea. Sound in sync was

used with both line standards and sub-carrier sound channels were also possible.

(5) Several design improvements were identified for the future and significant reduction in size appears to be possible. Compact low-power terminals for day-to-day short-distance operation with regional satellite systems would then seem to be practicable.

(6) Risk of interference to adjacent satellites appears to be manageable for the present in view of the limited number of satellites sharing the orbit and frequency band. Transmission standards suitable for occasional traffic can be achieved with relatively low power and high satellite gain. In due course improved energy dispersal techniques may be necessary and the most effective approach may eventually be the use of digital modulation.

(7) The World Administrative Radio Conference, WARC '79 introduced a modification to the international frequency table to permit land mobile-satellite systems in the 14–14·5 GHz up-link frequency band. This could be put to practical use for transportable television up-links in operational service with the European Communication Satellite, ECS, to be launched in 1982.

11 Acknowledgments

12 References

1 Neusten, M. and Marchant, B. 'Satellite relays and distribution', *IBA Technical Review*, No. 11, July 1978.
2 Salkeld, B., 'Energy dispersal for t.v. satellite up-links', IBC'80, IEE Conference Publication No. 191, pp. 309–12.

The Authors

Brian Salkeld (Member 1967), IBA Head of Network Planning since 1974, started his career in armaments research for the Ministry of Supply and then worked on missile guidance systems with Plessey. He worked with STC on microwave systems in a number of countries overseas before joining IBA, where his early work centered on development of television re-broadcast receivers. Mr Salkeld is currently a member of UK CCIR Study Group 4, CMTT and the EBU sub-group T7 studying the European Satellite System ECS in relation to Eurovision.

Sujay Verma (Member 1976) received the honours degree in electrical and electronic engineering from University of Wales in 1969. He joined Rediffusion (Wales) in the same year, where he was concerned with the planning of wired television networks. In 1971 he joined C & S Antennas where he designed and developed u.h.f. antenna systems. He joined the IBA in 1972, working first on the planning of u.h.f. television relay stations. He is at present a Senior Engineer in the Network Planning Section, working on microwave link systems for television and stereo radio.

David Griffiths joined the BBC in 1966 as a graduate trainee at the Daventry transmitting station. After a brief period investigating compact u.h.f. transmitting aerials he moved to the Studio Planning and Installation Department to work on the provision of video tape recording facilities at the BBC Television Centre. In 1968 he transferred to the Transmitter Planning and Installation Department, working on microwave link systems for television and stereo radio. He joined the IBA in 1975 to work on the expansion of television and radio link networks. He has been involved in television experiments with the European test satellite OTS since its launch in 1978 and is a member of the relevant EBU working group. His position is now Principal Engineer–Network Development in the Network and Planning Department.

Simple method for analysis of dielectric-coated wire antennas

Professor B. D. POPOVIĆ, D.Sc.*
A. R. DJORDJEVIĆ, D.Sc.*
and
N. M. KIRĆANSKI, M.Sc.*

SUMMARY

A simple method is presented for analysis of wire antennas assembled from arbitrarily interconnected straight-wire segments, which are completely or partially covered with a dielectric coating. The thickness of the coating is assumed to be not larger than about the radius of the wire, and its relative permittivity not larger than 10. The method is based on the so-called two-potential integro-differential equation combined with polynomial approximation of current. Numerical results are found to be in very good agreement with experiments.

* *Department of Electrical Engineering, University of Belgrade, PO Box 816, 11001 Belgrade, Yugoslavia.*

1 Introduction

Dielectric-coated wire antennas are used in many applications where direct contact of the metallic wire with the surrounding medium is undesirable. For example, the surrounding medium may be corrosive and/or conductive, and the dielectric coating may be protective and/or aimed at improving the antenna radiation properties. The thickness of the coating, in order to be efficient as a protection, is usually of the order of the wire radius. Such coatings influence the properties of the antenna to such an extent that bare-antenna theory cannot be applied for the antenna analysis.

There are only a few papers dealing with this interesting topic. A paper by King *et al.*[1] describes a very simple method for analysis of dielectric-coated cylindrical dipole-antennas driven by a delta-function generator. Although it yields accurate current distribution, the values of the antenna driving-point admittance obtained by the method cannot be considered sufficiently accurate. A paper by Richmond and Newman[2] presents a method comparable in generality with that proposed in this paper. However, it appears to be more complicated, because it is based on the reaction integral which is solved by piecewise-sinusoidal approximation for current. In addition, it also uses delta-function generator model, so that values of the antenna susceptance should be considered only as approximate. Recently two papers by James and Henderson[3,4] appeared, dealing with cylindrical antennas covered with a layer of dielectric or ferrite material. Their method is based on cavity-type modes excited in the coating and the variational method. However, it was applied only to cylindrical structures and is too complex when only thin coatings are considered.

This paper is aimed at presenting a novel simple method for analysis of antennas in the form of arbitrarily interconnected straight-wire segments, covered completely or partially with a dielectric coating. Essentially, it represents an extension of the method for analysis of such bare-wire structures.[5] It is based on the so-called two-potential integro-differential equation for current distribution, which is approximated by simple polynomials.[6] Coaxial-line excitation of the antenna is approximated by magnetic-current frill.[7] Very good agreement is observed between numerical results according to the proposed theory and available experimental results (obtained partly by the authors).

2 Outline of the Method

Consider a straight-wire segment of radius a and length $(s_2 - s_1)$, situated in a vacuum, covered with a dielectric layer of permittivity ε, as shown in Fig. 1. Let the complex current in the wire be $I(s)$, and its angular frequency ω. We wish to determine the field due to the

current and the polarized dielectric layer first on the segment axis, for later application of extended boundary condition, and secondly at relatively large distances from the wire, for later computation of the influence of one wire segment on another.

The wire being assumed a perfect conductor, there exists only the normal component of the electric field vector $\mathbf{E}$ on its surface. Provided that neither the thickness $d = b - a$ of the dielectric coating, nor its permittivity, ε, are large, we may assume that, approximately, the coating is polarized only radially. This results in two layers of polarization charges on the two cylindrical surfaces of the coating, of equal magnitudes per unit length and opposite signs. Radial polarization currents in the coating exist in addition to these charges. The influence of the coating on the field can very nearly be attributed to these sources.

The normal component of the electric field on the wire surface can be expressed in terms of the wire current, as follows. By continuity equation, surface charge density on the wire is given by

$$\sigma = -\frac{1}{j\omega 2\pi a}\frac{dI}{ds}, \qquad (1)$$

and the electric displacement vector intensity is simply $D = \sigma$. Thus $E(a) = \sigma/\varepsilon$, and intensity of the polarization vector

$$P(a) = (\varepsilon - \varepsilon_0)E(a) = \sigma(\varepsilon - \varepsilon_0)/\varepsilon.$$

For $a \leqslant R \leqslant b$ quasi-static approximation yields
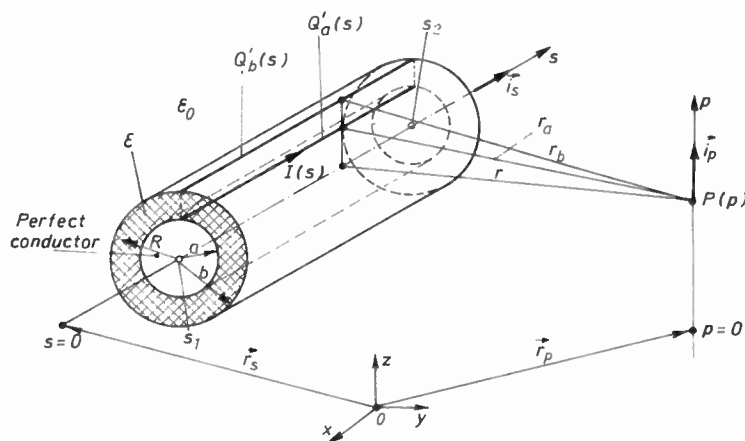
$$E(R) = aE(a)/R,$$

so that

$$P(R) = -\frac{\varepsilon - \varepsilon_0}{\varepsilon}\frac{1}{j\omega 2\pi R}\frac{dI}{ds} \quad (a \leqslant R \leqslant b). \qquad (2)$$

Polarization charge per unit length of the segment is thus

$$Q_a'(s) = -2\pi a P(a), \quad Q_b'(s) = 2\pi b P(b) = -Q_a'(s) \quad (3)$$

and the radial polarization current density

$$J_p(R) = j\omega P(R).$$

According to the extended boundary condition[8] we shall need the axial field produced by $I(s)$, $Q_a'(s)$, $Q_b'(s)$ and $J_p(R)$ along the segment axis. Radial polarization currents produce zero field at these points. Since axial symmetry of current $I(s)$ and charges $Q_a'(s)$ and $Q_b'(s)$ is assumed with respect to the segment axis (s-axis), the axial electric field on the axis is the same as if the total current and charges were located along lines at $R = a$ and $R = b$, as shown in Fig. 1.

When computing the p-component of the field due to these sources at a point P outside the segment, the same approximation can be used. To average the contribution of current and charges around the cylinders $R = a$ and $R = b$, it is obviously convenient to assume equivalent line current and charges along the lines of intersection of a plane perpendicular to r and s-axis with the two cylinders, as shown in Fig. 1.

Starting from the expression

$$\mathbf{E} = -j\omega\mathbf{A} - \text{grad } V, \qquad (4)$$

where $\mathbf{A}$ is the retarded magnetic vector-potential and V the retarded electric scalar-potential, it is not difficult to show that the p-component of the electric field vector of these line sources at a point P is given by

$$E_p = -j\omega\frac{\mu_0}{4\pi}\int_{s_1}^{s_2}\left[\left(\mathbf{i}_p\cdot\mathbf{i}_sI(s) + \right.\right.$$
$$\left. + \frac{\varepsilon_0}{\varepsilon k^2}\frac{dI}{ds}\frac{\partial}{\partial p}\right)\frac{\exp(-jkr_a)}{r_a} +$$
$$\left. + \frac{\varepsilon - \varepsilon_0}{\varepsilon k^2}\frac{dI}{ds}\frac{\partial}{\partial p}\frac{\exp(-jkr_b)}{r_b}\right]ds, \qquad (5)$$

where

$$k = \omega\sqrt{\varepsilon_0\mu_0}$$

is the free-space propagation coefficient,

$$r_a = (r^2 + a^2)^{1/2}, \qquad r_b = (r^2 + b^2)^{1/2} \qquad (6)$$

and

$$r = |\mathbf{r}_p + p\mathbf{i}_p - \mathbf{r}_s - s\mathbf{i}_s|. \qquad (7)$$

Fig. 1. A current-carrying wire segment covered with a dielectric layer.

Details of the derivation of eqn. (5) can be found in Reference 5, noting that in addition to scalar-potential V due to free charges on the wire segment surface there are polarization charges $Q'_a(s)$ and $Q'_b(s)$. Obviously, if the s- and p-axis coincide, then $r = 0$, and eqn. (5) yields the axial component of the electric field on the segment axis.

Consider now an antenna consisting of N arbitrarily interconnected straight-wire segments, such that each is covered completely with a dielectric coating of constant thickness, or is not covered at all. Let E_i be the impressed electric field on the wires. From eqn. (5) and the concept of extended boundary condition, one easily obtains that currents along the segments satisfy the following integro-differential equation (the so-called two-potential equation):

$$\sum_{i=1}^{N} \int_{s_{1i}}^{s_{2i}} \left[\left(\mathbf{i}_p \cdot \mathbf{i}_{s_i} I_i(s_i) + \frac{\varepsilon_0}{\varepsilon_i k^2} \frac{dI_i}{ds_i} \dddot{p} \right) \frac{\exp(-jkr_{ai})}{r_{ai}} + \frac{\varepsilon_i - \varepsilon_0}{\varepsilon_i k^2} \frac{dI_i}{ds_i} \dddot{p} \frac{\exp(-jkr_{bi})}{r_{bi}} \right] ds_i = -j \frac{4\pi}{\omega\mu_0} \mathbf{i}_p \cdot \mathbf{E}_i. \quad (8)$$

In this equation, ε_i is the permittivity of the coating of the segment i, and a_i and b_i the inner and outer radii of the coatings.

Solution of eqn. (8) can be obtained only numerically, for example by the point-matching technique. The details of the solution and of treatment of the junctions is exactly the same as in Reference 5, and will not be repeated here. We just note that all the results presented below were obtained using polynomial approximation of current along the segments,[6]

$$I_i(s_i) = \sum_{m=0}^{n_i} a_{im} s_i^m, \qquad i = 1, 2, \ldots, N, \quad (9)$$

and approximating coaxial-line excitation by equivalent frill of impressed magnetic currents, the electric field $\mathbf{E}_i$ of which can be determined relatively easily.[7]

3 Numerical Results

The theory outlined above was applied to a large number of cases for which experimental results were either available in the literature, or were obtained by the authors. In all the cases considered, agreement between theoretical and experimental results was found to be very good. Some examples are presented below.

The theory was first checked by comparing the theoretical results with experimental results of Lamensdorf.[9] In that paper, values of measured admittances and current distributions were presented for monopole antennas of radius $a = 3.18$ mm, representing a simple protrusion of the inner conductor of a coaxial line of outer conductor radius $c = 9.53$ mm. A long dielectric tube of inner radius equal to the radius of the monopole was pressed over the monopole down to the ground plane. The thicknesses $d = b - a$ of the dielectric coating thus obtained were 3.18 mm, 8.71 mm and
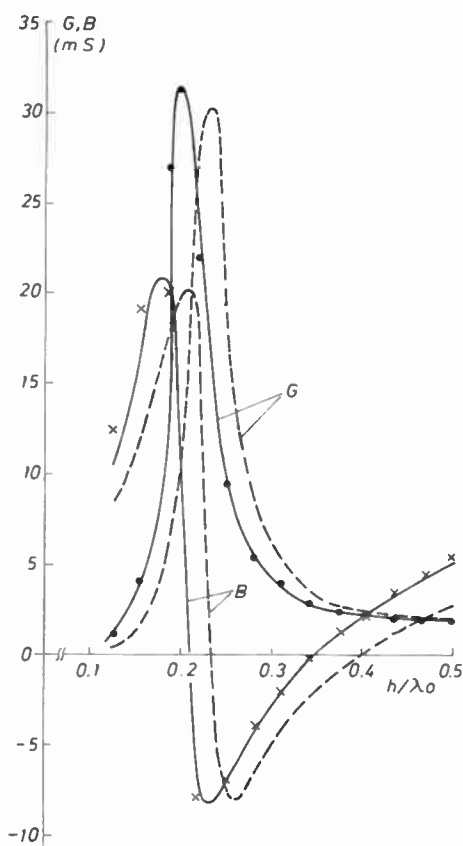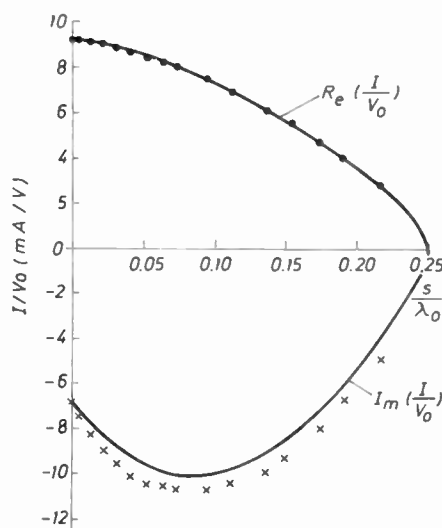

Fig. 2. Conductance (G) and susceptance (B) of a bare and a dielectric-coated monopole antenna versus the normalized antenna length h/λ_0 for $\lambda_0 = 0.5$ m, wire radius 3.18 mm, dielectric layer thickness 3.18 mm and relative permittivity 3.2.
———— theory, with coating.
— — — theory, without coating.
● ● ● experiment, with coating.[9]
× × ×

Fig. 3. Current distribution along a dielectric-coated monopole antenna for $h/\lambda_0 = 0.25$. Other data are the same as for Fig. 2.
———— theory.
● ● ● experiment.[9]
× × ×

Fig. 4. Examples of dielectric-coated wire antennas: (a) vertical monopole, (b) inclined monpole and (c) monopole-annenna with a parasitic element. All dimensions are in millimetres, wire radius is 3 mm, dielectric layer thickness 2 mm and relative permittivity 2·1. Not drawn to scale.
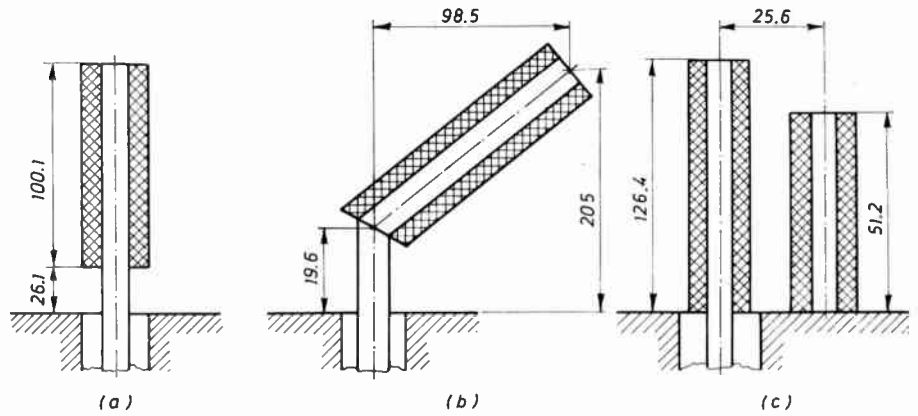
22·23 mm, and relative permittivities of the coating were $\varepsilon_r = \varepsilon/\varepsilon_0 = 3\cdot2$, 9 and 15. Frequency in all the measurements was $f = 600$ MHz. For $\varepsilon_r = 3\cdot2$ and $d = 3\cdot18$ mm and 8·71 mm agreement between theoretical and experimental results was found to be excellent. For $\varepsilon_r = 9$ and $d = 3\cdot18$ agreement was still good, but for larger values of ε_r and d discrepancies were progressively larger. As an example, Fig. 2 shows

conductance and susceptance for a bare and a dielectric coated antenna, versus normalized antenna length h/λ_0 (λ_0 is the wavelength in free space). Shown in Fig. 3 is theoretical and experimental current distribution for such an antenna of length $h/\lambda_0 = 0\cdot25$. Very good agreement between theoretical and experimental results is observed, as well as a significant influence of the coating on the antenna parameters.

To verify the theory on more complex structures, the authors constructed a number of antennas of wire radii $a = 3$ mm, covered partially with teflon coating ($\varepsilon_r = 2\cdot1$) of thickness $d = b - a = 2$ mm. Excitation of antennas was performed by means of a 50 Ω coaxial line of outer conductor radius $c = 6\cdot9$ mm, the first antenna
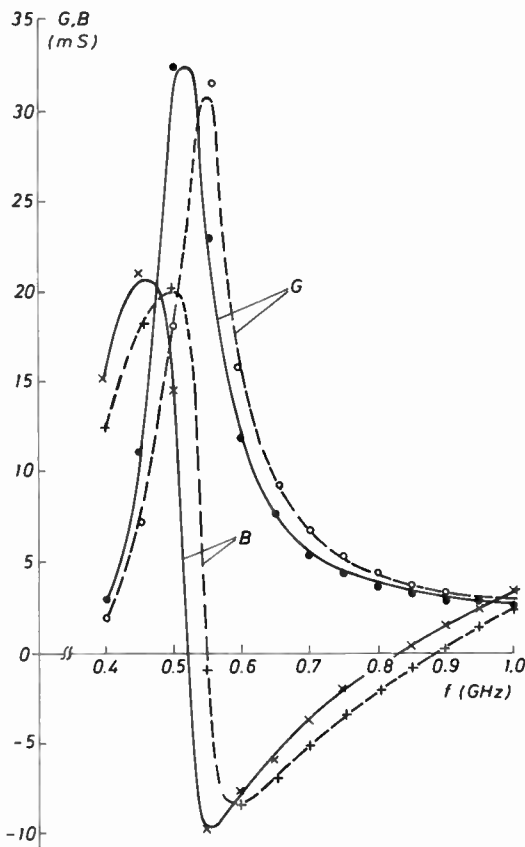

Fig. 5. Conductance (G) and susceptance (B) of the monopole antenna shown in Fig. 4(a), versus frequency.
——— theory, with coating.
— — — theory, without coating.
● ● ● experiment, with coating.
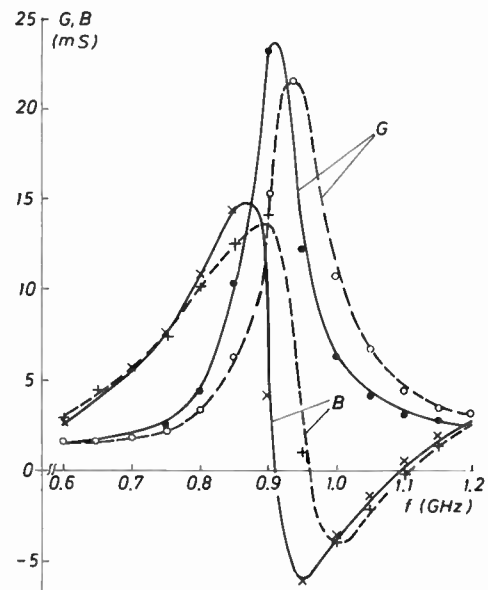× × ×
○ ○ ○ experiment, without coating.
+ + +

Fig. 6. Conductance (G) and susceptance (B) of the inclined monopole antenna shown in Fig. 4(b), versus frequency.
——— theory, with coating.
— — — theory, without coating.
● ● ● experiment, with coating.
× × ×
○ ○ ○ experiment, without coating.
+ + +
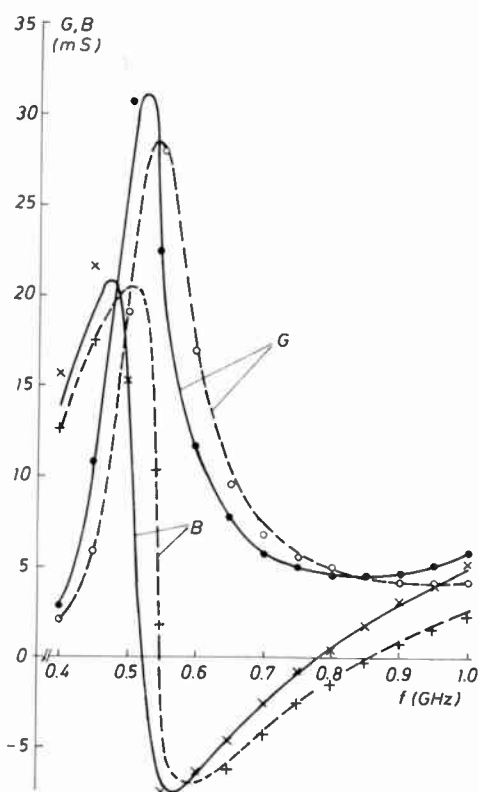
Fig. 7. Conductance (G) and susceptance (B) of the monopole-antenna with a parasitic element shown in Fig. 4(c), versus frequency.

——— theory, with coating.

— — — theory, without coating.

● ● ●
        experiment, with coating.
× × ×

○ ○ ○
        experiment, without coating.
+ + +

segment being always a simple protrusion of the inner coaxial-line conductor ($a = 3$ mm). Shown in Fig. 4 are some of the structures analysed experimentally, for which the results are presented below.

Figure 5 shows conductance and susceptance of the monopole antenna shown in Fig. 4(a) versus frequency, together with the results for the bare antenna. Figure 6 shows analogous results for the inclined monopole antenna of Fig. 4(b), and Fig. 7 for the dielectric-coated antenna with a parasitic element, shown in Fig. 4(c). Very good agreement between theoretical and experimental results is observed in all the cases.

A general conclusion from all the examples is that the coating results in antennas having slightly sharper G and

B curves near resonance than bare antennas. Also, it is seen that coatings of even moderate thicknesses and values of relative permittivities have strong influence on antenna properties. Bare-antenna theory cannot be applied for obtaining reasonably accurate results in these cases.

4 Conclusions

A simple method is presented for numerical analysis of wire antennas assembled from arbitrarily interconnected straight-wire segments with relatively thin dielectric coatings. Coaxial-line feed is modelled by a frill of impressed magnetic currents. Theoretical results obtained according to the proposed theory are in very good agreement with experimental results, provided that thickness of the coating is not larger than about radius of the wire, and its relative permittivity less than about 10.

Bare-antenna theory cannot be applied for even moderately accurate analysis of such antennas. General influence of the coating is to sharpen the G and B curves near resonance, and (as expected) to shift these curves towards lower frequencies.

5 References

1 King, R. P., Lee, K. M., Mishra, S. H. and Smith, G. S., 'Insulated linear antenna: theory and experiment', J. Appl. Phys., 45, p. 1688, 1974.

2 Richmond, J. H. and Newman, E. H., 'Dielectric coated wire antennas', Radio Sci., 11, no. 1, p. 13, January 1976.

3 James, J. R. and Henderson, A., 'Electrically short monopole antennas with dielectric or ferrite coating', Proc. Instn Elect. Engrs, 125, no. 9, p. 793, September 1978.

4 James, J. R. and Henderson, A., 'Investigation of electrically small VHF and HF cavity-type antennas', Proceedings of IEE International Conference on Antennas and Propagation, London, p. 322, November 1978.

5 Djordjević, A. R., Popović, B. D. and Dragović, M. B., 'A method for analysis of wire-antenna structures', Archiv für Elektrotechnik, 61, p. 17, 1979.

6 Popović, B. D., 'Polynomial approximation of current along thin symmetrical dipoles', Proc. Instn Elect. Engrs, 117, no. 5, p. 873, May 1970.

7 Tsai, L. L., 'A numerical solution for the near and far fields of an annular ring of magnetic current', IEEE Trans. on Antennas and Propagation, AP-20, no. 5, p. 569, September 1972.

8 Waterman, P. C., 'New formulation of acoustic scattering', J. Acoust. Soc. Am., 45, p. 1417, June 1969.

9 Lamensdorf, D., 'An experimental investigation of dielectric-coated antennas', IEEE Trans. on Antennas and Propagation, AP-15, no. 6, p. 767, November 1967.

Contributors to this Issue

Ken Baker has been with the Plessey Company for 11 years, the first five of which were spent at the Caswell Research Laboratories designing silicon integrated circuits. He then worked for two years as Product Manager for Telecommunication I.C.s with Plessey Semiconductors and joined Plessey Microsystems in 1977. He is currently employed as Marketing Manager for memory products and is responsible for product planning the commercial aspects of core, semiconductor and magnetic bubble memory systems. Mr Baker studied at Lanchester Polytechnic and Leicester University and has been a chartered engineer since 1969. He is currently a member of the Institution's Components and Circuits Committee and of the Papers Committee. He is the author of eighteen published articles for conferences and in the technical press.

Nicholas David received the Engineering Diploma in electrical engineering from the University of Patras, in 1976 and the M. Eng. in electrical engineering from Carleton University, Ottawa, in 1978. Since 1978 he has been with the Electro-magnetics Laboratory at the University of Patras working as a teaching assistant and studying toward the Ph.D. degree in digital communication systems. He has published several papers in the field of digital modulation.

Vassilios Makios received his Dipl. Ing. in electrical engineering in 1962 from the Technical University in Munich, and his Ph.D. in 1966 from the Max Planck Institute for Plasmaphysics and the Technical University in Munich; from 1962 to 1967 he worked as a Research Associate in the Institute. In 1967 Dr Makios went to the Department of Electronics, Carleton University, Ottawa, and successively held the appointments of Assistant Professor, Associate Professor and Professor, and he is now Adjunct Professor. Since 1974 he has occupied the Chair of Electromagnetics and is the Director of the Laboratory of Electromagnetics in the Electrical Engineering Department of the School of Engineering of the University of Patras.

Professor Makios has worked in shock wave plasma, microwave diagnostics, microwaves, communications electronics, radar remote sensing, CO_2 high pressure laser research and integrated optics. He is currently working on microwaves, communications electronics and radar applications. He has published over 60 papers in scientific journals and holds three patents. A Consultant to many Canadian and Greek organizations, he is the Greek National Delegate to AGARD.

Branko Popovic received the B.Sc., M.Sc. and D.Sc. degrees of the University of Belgrade in 1958, 1963 and 1967 respectively. From 1959 to 1966 he was employed as an 'Assistant', and in 1966 he was appointed a 'Docent' with the Department of Electrical Engineering of the University of Belgrade. From 1971 to 1977 he was an Associate Professor, and since 1977 he has been a Professor with the Department of Electrical Engineering of the University of Belgrade.

From January 1968 to September 1969, Professor Popovic conducted graduate and under-graduate courses at the Electrical Engineering Department of Virginia Polytechnic Institute, Blacksburg, Virginia, as an exchange visitor, and during the spring semester of 1973 he held a Visiting Associate Professorship at McGill University, Montreal, Canada. He has been a consultant to the Institute Nikola Tesla in Belgrade, doing research on submarine power cables, and to the Radio Industry Zagreb, participating in the design of medium-wave antenna arrays.

Professor Popovic has twice been awarded the Distinguished Teachers' Award by his Belgrade students, and he was a co-recipient of the Belgrade October Prize for technical sciences in 1967. In 1978 he was elected a Corresponding Member of the Serbian Academy of Science and Arts. He is the author or co-author of numerous technical papers and has written two textbooks. One of his papers to this Journal gained him the Heinrich Hertz Premium for 1973.

Antonije Djordjević received the B.Sc., M.Sc. and D.Sc degrees from the University of Belgrade in 1975, 1977 and 1979 respectively. During his studies he won several prizes in national and international competitions in physics, mathematics and electrical engineering and was awarded the October Prize for his Diploma work. Dr Djordjević is at present a teaching assistant with the Department of Electrical Engineering, University of Belgrade, and is mainly concerned with antenna theory and techniques.

Nenad Kirćanski received his B.Sc. and M.Sc. degrees from the University of Belgrade in 1977 and 1980 respectively, being awarded the October Prize for his Diploma work. In 1977 he joined the Institute Mihailo Pupin where he worked in the field of robotics. Since 1980 he has been a teaching assistant with the Department of Electrical Engineering, University of Belgrade, and is a member of the antenna group.