THE UBIQUITOUS TRANSISTOR

INAUGURAL LECTURE

Delivered at the
University of Wales, SWANSEA
on 9th May, 1988

by

Kenneth Board

Professor of Electronic and Electrical Engineering,
UNIVERSITY of WALES, SWANSEA

UNIVERSITY of WALES, SWANSEA

First published 1988 by University of Wales, Swansea

Obtainable from Registrar's Office, University of Wales, Swansea, Singleton Park, Swansea, UK SA2 8PP

ISBN 0 86076 055 3

Printed by University of Wales, SWANSEA

1 INTRODUCTION

Perhaps I should start with a word of explanation about my title. The word 'transistor' appears there for two very simple reasons. First it is without question the single most important electronic building block and has played a central role in the period of rapid change we have gone through over the last decade. Secondly, it, together with some associated devices, has been the focus of my research interests since graduating here in 1965.

The presence of the work 'ubiquitous' needs a little more explanation. It is defined as 'being or seeming to be in many places at once'. As we shall see the transistor appears to satisfy this definition very well.

Let us start with the furthest-away example of the existence of transistors Pioneer 10 left the Solar System in June 1983 and is now some 7500 million miles from earth. It was launched in 1972 and would have many transistors aboard, functioning in its communication, guidance and navigation systems. Actually the launch date would enable us to pinpoint fairly accurately the vintage of transistor aboard.

In the solar system there are many examples where space probes have been sent to planets to gather data. Some have moved on into space, others crash landed on the plane surface.

Around the earth the frequency of occurrence increases rapidly. All the satellites presently in orbit carry large numbers of transistors in the on-board computer and communication systems.

And so we can proceed: In the air, all aircraft both civil and millitary, will carry many millions of transistors aboard, each functioning with extreme reliability and all following the general rule that the more recent the vintage, the smaller, faster and cheaper they would be.

At sea there is no large vessel without its navigation system, or on-board computer.

But it is on land that the real proliferation occurs. Transistors and transistor-based systems pervade the home, appearing in radio and television sets, in washing machines and video recorders, telephone systems and home computers. In industry, and in the office, where they occur in very large numbers indeed, they have a similar widespread influence.

Figure 1 shows the annual world-wide production of integrated circuits. With up to a million transistors formed on each chip the number of transistors being produced annually is around 10^{15} - with output increasing rapidly each year both because of the increasing demand for integrated circuits and the continuing increase in the number of transistors being integrated.

The range of size of transistors is shown in figure 2. The smallest size routinely seen in production consumes an area of about 10^{-6} cm^2 although in a research context the active region has been made very much smaller. The largest, shown in figure 3b consumes an area of 50 cm^2 and can control a current of 400 Amperes and withstand a voltage of 1500 volts.

Finally we should comment on the changes the transistor has brought about in the academic sector.

First it has radically changed teaching and research techniques in almost all disciplines through better apparatus, more powerful computers, computer based library systems and more general access to information through easier communication. Not entirely trivial are the much improved ways these days of preparing written documents through word-processing, and desk top publishing systems.

In teaching there is now common access to computers for each student that would have been unheard of a few decades ago. We all know I'm sure of the example of the present day PC. In the days before the transistor such a computer would have filled most of the floor-space in this auditorium, have dissipated hundreds of kW of heat, and in reliability terms have functioned on average for about 20 seconds between failures.

Apart from the effect on the teaching of traditional subjects, we have also seen the emergence of new subjects that would not have been possible at all. A good example is computer science, and we shall see later how advances in this subject have in a curious way, aided the development in turn of the transistor itself.

In summary, then, we can trace the origins of the so-called IT revolution to the invention of the transistor. It would in my view, be wrong to suggest

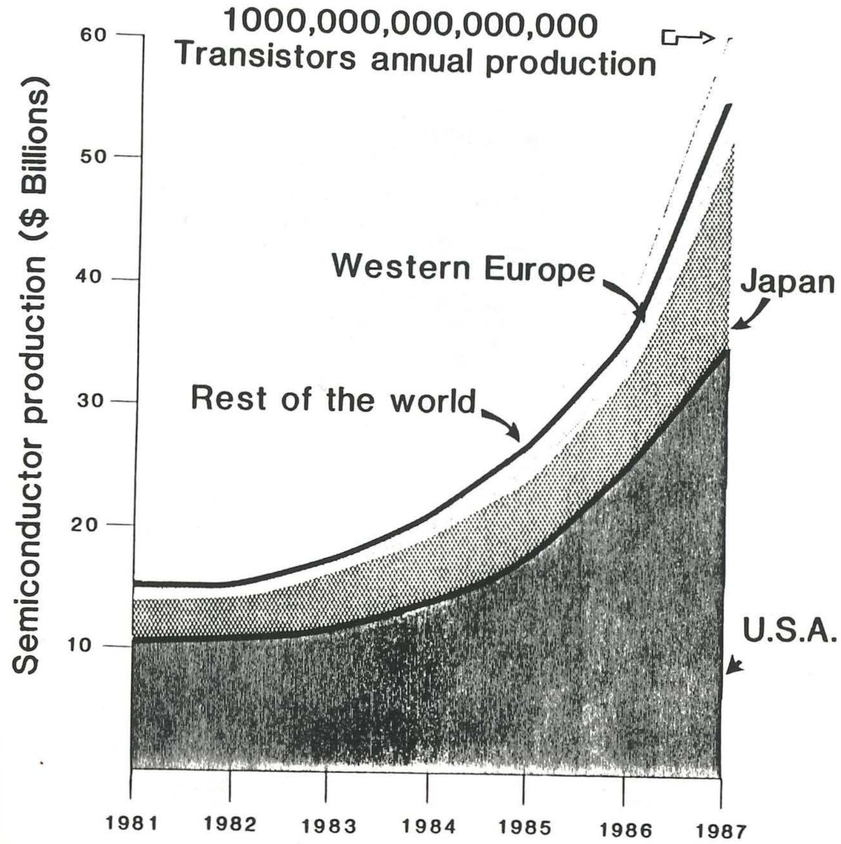


FIGURE 1

WORLD-WIDE PRODUCTION OF INTEGRATED CIRCUITS

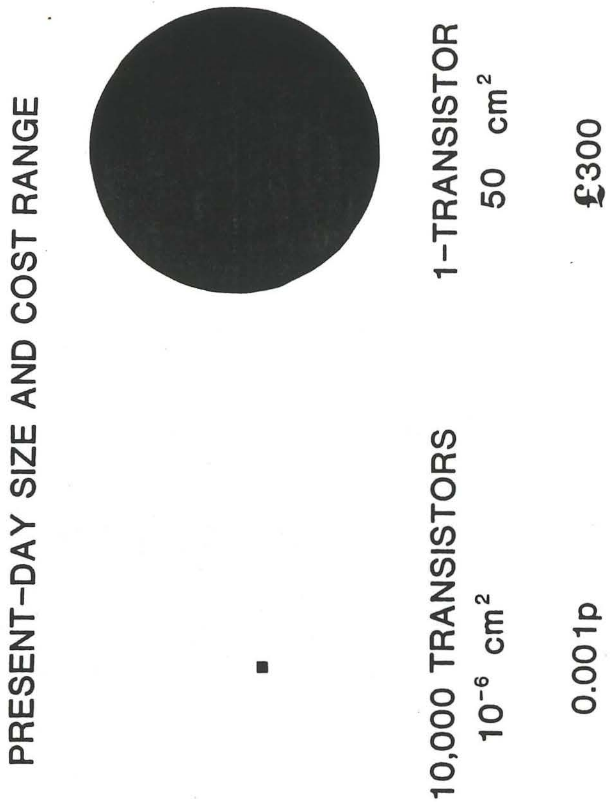


FIGURE 2

that it would not have happened at all without its invention. What is certain however, is that without it these changes would not have occurred when they did, nor on the scale they did. For the transistor has particular properties such as extremely high reliability and ease of cheap integration which were not appreciated fully by its inventors at the time, that has made all we have seen possible. Had some other option been taken at that time, then developments would have taken a different course and the period of rapid change we have witnessed perhaps delayed.

So what is a transistor, and how was it invented? This story is a fascinating one and I propose to spend a little time describing how it happened. Before doing so, however, I would like to identify two themes that will run through my talk and which will surface here and there as we proceed. The first theme is that of *Collaborative Research.* This is a fashionable topic nowadays but, as we shall see when we discuss the transistor is nothing new. The second theme is a more personal one and will touch upon some of my own involvement in the general area of semiconductor device research.

2 HISTORY OF THE TRANSISTOR

Ever since electrical energy has been used to perform useful functions it was considered desirable to control the flow of current without resorting to mechanical movement. After all currents were carried by electrons and it seemed a very crude way of starting and stopping the current flowing.

Interestingly this requirement cut right across the thinking of the day among engineers of other disciplines. When designers of electrical equipment attempted to join the engineering profession they were refused on the grounds that "it can't be engineering because there are no moving parts".

The quest for a purely electrical valve had earlier been satisfied by the vacuum tube. Sir William Crookes in the mid 19th century had been the earliest pioneer of this form of valve and Sir Alexander Fleming had produced the first crude form in 1904. This had no physical moving parts and could not only turn currents on and off, but could give amplification. However, it seemed somehow illogical to go to considerable lengths to get electrons out of the solid, control them and then get them back in again! Also the valve had a number of disadvantages which led researchers to look for an alternative. In particular it could never have served as a vehicle for the developments we have just illustrated because of its poor reliability, fragility, and cost.

Now the way in which this solid valve was perceived to work was rather like the way some civil engineers might control water flow. Their problem was how to produce a potential barrier in a solid, which is where we have our first example of interaction between disciplines. The developments in quantum mechanics by Sommerfeld Bloch and Wilson in the late 1920's suggested among other things, that such barriers could not exist in metals, but could in semiconductors such as germanium and silicon.

Well the upshot was the postulation by William Shockley of something which could control current and give amplification which came to be known as the transistor. Figure 3 shows Shockley's first notebook entry on the device - written on Christmas Eve in 1939.

Now it was 1947 before the transistor was actually realised, and this brings me the the next example of interdisciplinary activity. The evolution of events is illustrated in figure 4. The reasons for the original idea not working gave rise to fairly intensive research into semiconductor surfaces which led in the medium term to the demonstration of transistor action, but in its own right prompted a line of more basic research which is still ongoing.

An even more dramatic example of this kind of delay is that of the second major form of transistor the Field Effect Transistor.

The 'field effect' and its use as an amplifier was first proposed in 1930 by Lilienfield. However, it was first demonstrated 18 years later by Shockley in 1948. But it was in 1960 that the field-effect transistor was first demonstrated by Khang and Atalla.

These developments and their timescales are illustrated in figure 5.

This success was due, not only to the engineers whose motivation was the device itself or the physicists who provided the required physical background, but in no small part to the chemists and materials scientists who had had to make advances in the preparation of semiconductor crystals to the high degree of purity necessary, and the development of the technology for their fabrication. Thus it was very much a multidisciplinary effort that brought it about.

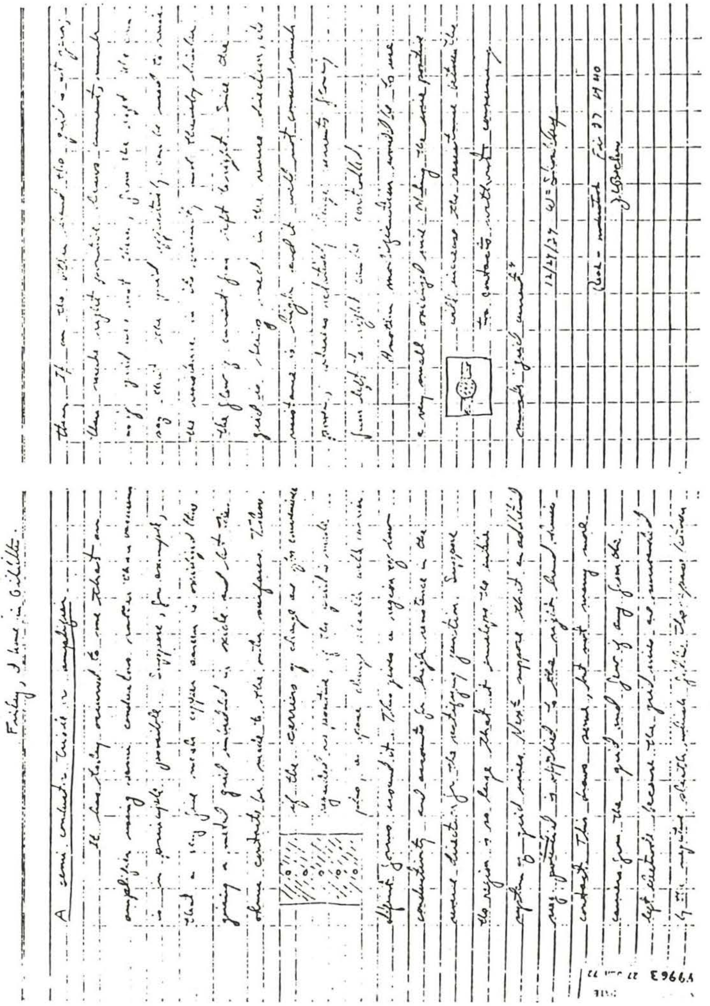


Figure 3 First notebook entries describing a workable transistor structure.

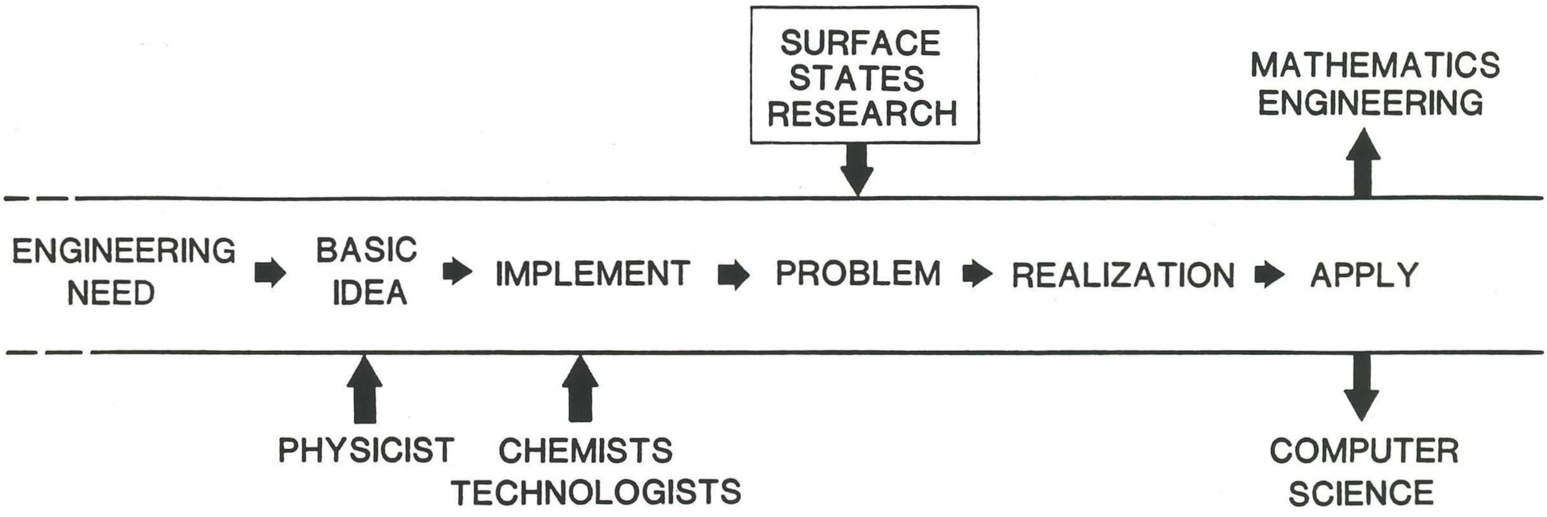


FIGURE 4

"MILESTONES IN INFORMATION TECHNOLOGY DEVELOPMENTS"

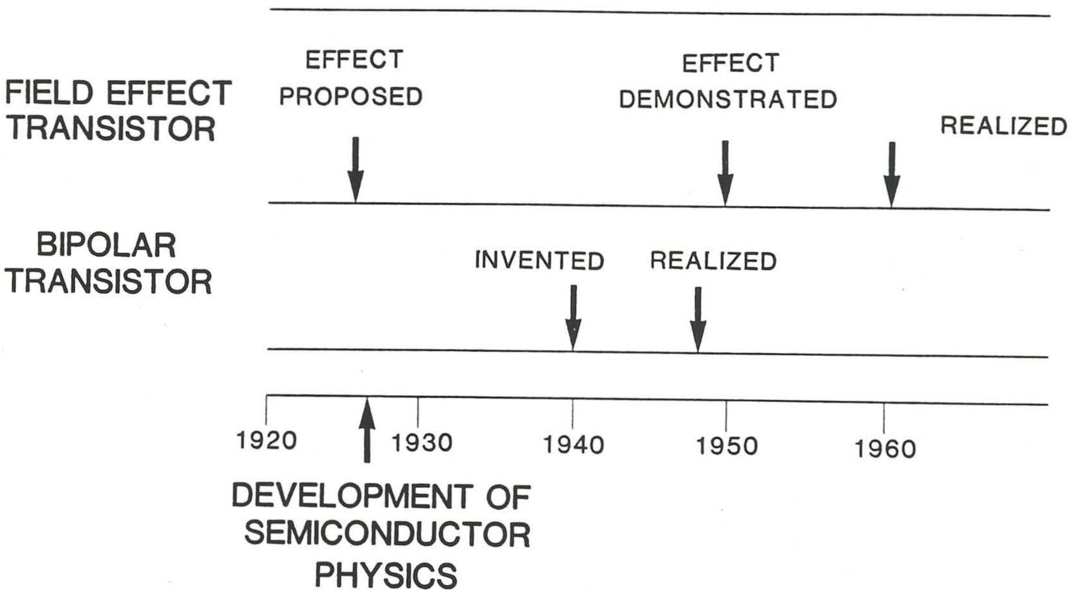


FIGURE 5

TIMESCALES OF TRANSISTOR DEVELOPMENT

3 DEVELOPMENT OF THE INTEGRATED CIRCUIT

The gap between the production of a single transistor in 1960 and the complex chip of today, carrying up to a million transistors on a piece of silicon not a great deal larger than the single transistor version is a profound one. The integrated circuit is not simply a large collection of transistors, but a complex system with all the difficulties in conception and design that that entails.

The incentive to take this route was provided among others by the military who had realised the benefits of communications, navigation and guidance systems which could only be realised electronically. The B29 World War II bomber had carried what was then the most complex electronic system yet devised - consisting of a thousand vacuum tubes. Can you imagine what a design achievement that was. A thousand glowing electronic valves installed in the noise-filled, vibrating environment of a bomber on active service with the vulnerability of the system to failure through gunfire etc. In those days I'm fairly sure that no redundancy was used in design, so that only one valve had to fail to render the system inoperative.

But the military had, despite such difficulties, caught a glimpse of the benefits of electronic 'integration' and proceeded to fund research projects with a view to achieving higher levels of component integration with improved reliability.

Perhaps surprisingly, although we today think that the extension from one transistor to many connected together on a similar-sized piece of silicon is an obvious one, it was certainly not the case at the time. Then people were dubious and thought that the way to go was to devise clever ways to connect together many single transistors reliably and cheaply but essentially using the copper wire and soldering techniques that were traditional.

The first suggestion to form circuits in silicon was made by Dummer of the Radar Research establishment in 1952. However, it was in 1959 that patents were filed in the USA staking claim to the invention of the integrated circuit. The scepticism with which this suggestion was greeted may be judged by the following comments from various organisations at the time:

1) The concept did not make optimum use of materials. Nichrome made better resistors and Mylar better capacitors. Performance of the transistors might be degraded by the inclusion of other components.

2) Circuits of this type were not producible. Component yields were always low, and if 20 components each with 90-percent yield were fabricated monolithically, the overall yield would be about 12 percent.

3) Designs would be expensive and difficult to change. Circuit designers would be out of a job.

The timing of events at this stage is noteworthy. The transistor, having been under development for more than thirty years following its conception, had depended on advances in quantum mechanics and their application to semiconductors. It was then available at the very time when pressures to produce complex electronic systems had also developed.

The way in which transistor integration has evolved is epitomised by Moore's Law, illustrated in figure 6. The exponential growth that has occurred since 1960 has shown some signs of levelling off but is still projected to reach 3 million components by 1992. We will have occasion to refer to this law a little later on.

4 DIFFICULTIES AND OPPORTUNITIES OF THE EXPANSION

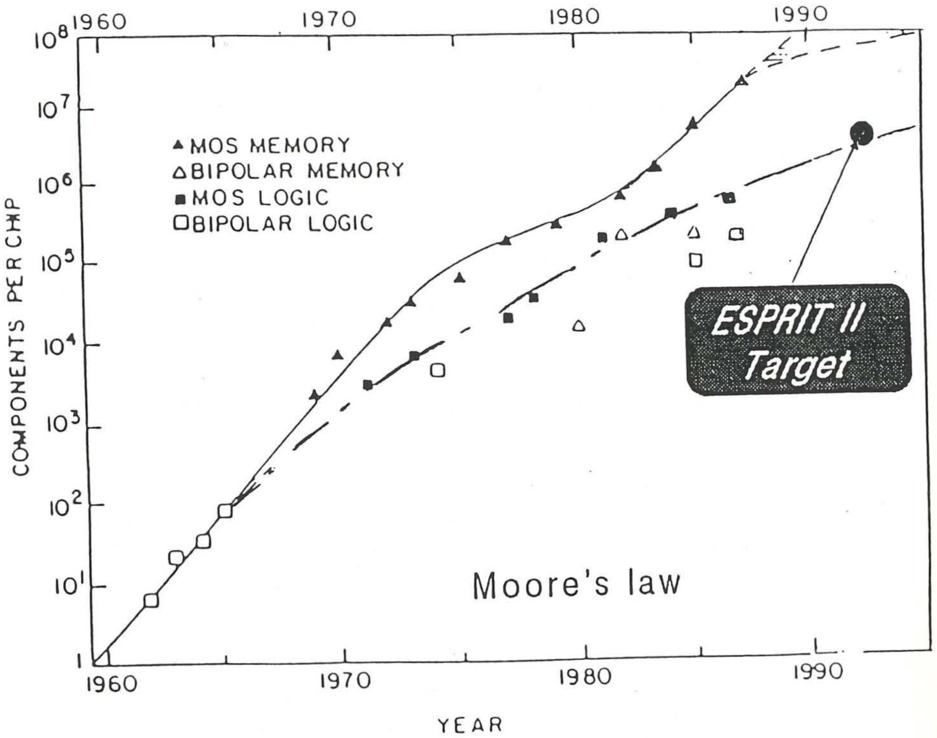
Generally when change occurs even in such diverse areas as population density, or the value of sterling, or commodity prices, it seems that it is not so much the magnitude of the change that causes difficulties, but how rapidly it occurs. In electronics the levels of integration have gone from one to one million in just thirty years. Such a rapid rate of change has inevitably given rise to problems which are many and varied. Some are moral, others philosophical, most difficult to solve. I want to confine my discussion to those which affect the academic. In particular I want to touch on the effect on teaching and research in turn.

A THE EFFECT ON TEACHING

The half-life of the electronic engineer is presently about five years. This means that half of what he is taught today as current technology will in five years be obsolete.

So what do we do? Do we concentrate on fundamentals and send out

High Density ICs



Complexity around
2 to 4 millions devices by 1992

FIGURE 6

graduates to industry who have little idea about current design strategies and technologies, or do we teach material that we know is going to become out of date?

Well, we could spend the rest of the talk debating this matter, but I want to concentrate my discussion on those problems that relate to research.

However, before leaving this topic, it seems to me that however we view these difficulties it becomes increasingly important to develop techniques for self-updating, and to communicate these to students (and, of course, staff).

Of course, it would be wrong to emphasize such difficulties without also referring to the opportunities in teaching that the advances in technology have given directly through computer assisted learning and indirectly through improved laboratory facilities and equipment. Indeed, it may well be that the use of these tools will provide the best way to cope with these difficulties.

B THE EFFECT ON RESEARCH

And now to discuss some of the difficulties and opportunities of the revolution that relate to research. This I want to illustrate by way of an example relating to some device research I was personally involved in.

In 1981 I was involved at Cornell University in the development of a new rectifying structure in another type of semiconductor-gallium arsenide. On this occasion technology developments had preceded the ideas we had about device structures. A new technology called Molecular Beam Epitaxy had emerged and this allowed us to control impurities within a semiconductor in a manner unheard of before.

I don't wish to go into too much detail about the new device. Essentially there had historically been two forms of rectifying structure in general use - the metal-semiconductor junction and the p-n junction. The device we proposed combined many of the positive attributes of both but was free from several of their most serious disadvantages.

Well the device was conceived, a first order theory developed and test structures designed and fabricated. They gave results which agreed with our main prediction as shown in figure 7.

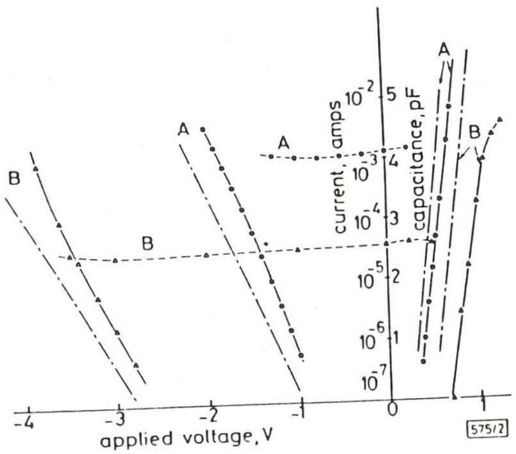


Figure 7 *Current and capacitance against voltage of p.d.b. device*

There were several features of that work that were particularly satisfying to me.

First and foremost it worked out to be self-contained. The predictions made had been borne out and we had been the first to propose and demonstrate it.

Secondly, it has led to a considerable number of other researchers taking up work on its theory and application. In particular, it has been applied to new forms of transistor.

Thirdly, one of the applications we suggested for it then - that of a high-speed balanced mixer - has recently appeared as a commercial device in the Hewlett Packard catalogue.

Finally, it led us to conceive a new form of fast optical switch which was demonstrated the following year at Cornell. On this last device, we hold the UK and US patent rights.

Now all this was very stimulating and challenging. However, the work was only made possible by access to equipment that was far too expensive for us to contemplate having it at Swansea. Today's cost of an MBE system is close to £1M and that excludes support staff, maintenance and ancillary equipment. It was therefore impossible to contemplate setting up such an activity here in Swansea. At the time in 1981, there were very few such machines in the UK - and certainly none in the University System. If and when such an event occurred it was highly likely that it would be placed in one of the so-called Central Facilities - Edinburgh, Southampton or Sheffield.

Well, a change of direction was called for. At that time in the early 1980's, it was perceived in Europe that despite the usual significant contributions by Europeans to the formative stages of the IT revolution, it was falling well behind in the application and exploitation of the technology.

Thus the EEC launched its 10 year shared-cost programme, spread over the period 1984 to 1993. It occurred in two phases, the first running from 1984 to 1989 at a level of approximately 2000 Mecu and the second projected to run until 1993. The funding areas largely reflect those in which very rapid developments are occurring and cover Information Technology, Advanced Industrial Technologies and Alternative Energy Sources.

The transistor and its integration on a large scale has had a major influence on each of these, although perhaps not as large in the Alternative Energy and Industrial Technologies Sectors as in the IT programme, ESPRIT.

The projected funding for the second phase of ESPRIT is substantially larger than phase 1 although having gone through the mincing machine of the European summit, with arguments about the CAP etc., the figure for IT now stands at 1,6 Mecu instead of the originally proposed value of 2.05 Mecu.

The particular programme we are concerned with at Swansea is ESPRIT.

5. THE ESPRIT PROGRAMME

ESPRIT is a shared-cost programme, largely controlled by the so-called "Big-Twelve". These are the 12 major electrical/electronic companies in Europe and include GEC, Plessey, ICL from the UK, and Philips, Siemens and several others from Europe.

Its main objectives are:

- to provide European IT Industry with the basic technologies it needs to meet competitive requirements of the nineties.
- to promote European Industrial Cooperation in IT.
- to contribute to the development of internationally accepted standards.

Now the work involved has to come somewhere between the more basic research which is funded separately and that which is too close to the marketplace, and would cause problems for industrial partners who may be in competition in certain product areas.

At this point I would like to switch from generalities and discuss the particular part of ESPRIT which we in Civil and Electrical Engineering are concerned with and how we come to be involved.

Here I think you will again see clearly the theme of interdisciplinary and collaborative activity which I regard as so important.

The main sections covered by the ESPRIT programme are 1) Scope, 2) Microelectronics, 3) Information Processing, and 4) Applications.

The one with which we are concerned is Microelectronics. Within microelectronics, three sections are identified. They are:

1 High Density Integrated Circuits

2 High Speed Integrated Circuits

3 Multifunction Integrated Circuits

It is in the first (High Density Integrated Circuits) that our programme exists.

Professor Morris in his recent inaugural referred to "Right first time" design. Nowhere is this more important today then in the design of advanced silicon circuits.

In the old days circuits containing tens or hundreds of transistors could be bread-boarded, tested, and modified several times until the specification was met. Nowadays the most advanced high density circuits in silicon can consume up to a 100 man-years of design effort. A present rule of thumb is that a man-year costs around £70 K, so that the design cost is near £7 M. As important is the lead time involved, since a company getting to market only weeks before another can make a major difference to profitability. Thus effective design tools are a vital necessity.

This need was perceived in 1982 when Professor Owen and myself discussed the possibilities of applying the Finite Element technique to the design of transistors together with two industrial partners, Philips and GEC, the Rutherford Laboratory, and another University partner (Trinity College, Dublin). Together we obtained European Funding in a scheme which preceded the main ESPRIT programme. Again the collaborative element was strong. Here we were, Civil and Electronic Engineers from Swansea, getting into partnership with mathematicians and physicists from the other sites. The start was halting. I well remember hearing continually from the Civil Engineers about something called a stiffness Matrix. We well knew what a capacitance Matrix was, but it was some time before we realised they were the same thing in mathematical terms.

Another memory I have from the start of this collaboration was the attitude of one of the industrial partners, who shall, be nameless, which can be summed up as "we are in this for the money, we are so far ahead of you lot that you can't teach us anything, and we are going to do nothing to help you catch up. Well we were behind in those days, after all, we were just starting in a new field while they had already committed something like 12 man-years to it.

Five years on things have changed a lot for the better. Within a year we had a 2-D off-state computer code working, which is in use as a design tool in industry - and I'll be giving an example of what it can do later on. Also that earlier EEC project was succeeded in 1986 by a type A ESPRIT project in which the number of partners has grown to ten, including two Italian, one Belgian and two further Irish partners. The project also includes three of the so-called "Big-Twelve" partners.

We have just gone through a very successful mid-term review on that project and a measure of how attitudes have changed can be seen by a remark made by the industrial partner referred to earlier, in response to a comment from us about the state of readiness of our major deliverable. "What you are doing will be valuable to us and you'll probably get there first".

And now to illustrate some of the fruits of this work.

6. EXAMPLES IN TRANSISTOR MODELLING

A Power DMOS

The Wolfson Foundation award, made in 1981, was concerned with power transistors and their modelling. Transistors typically used in computer systems are very small and operate at 5-15 volts only. The power transistor in question was designed within our microelectronics group with the mask layouts done on Siliconix's Applicon system. It was designed to operate at 1200 volts - a voltage which opens up possibilities of mains operation and on-chip integration with low-voltage logic circuitry.

A schematic diagram of the active part of the transistor is shown in figure 8 and the resulting potential and field plots shown in figures 9 and 10.

The high field regions show where avalanche breakdown is likely to occur.

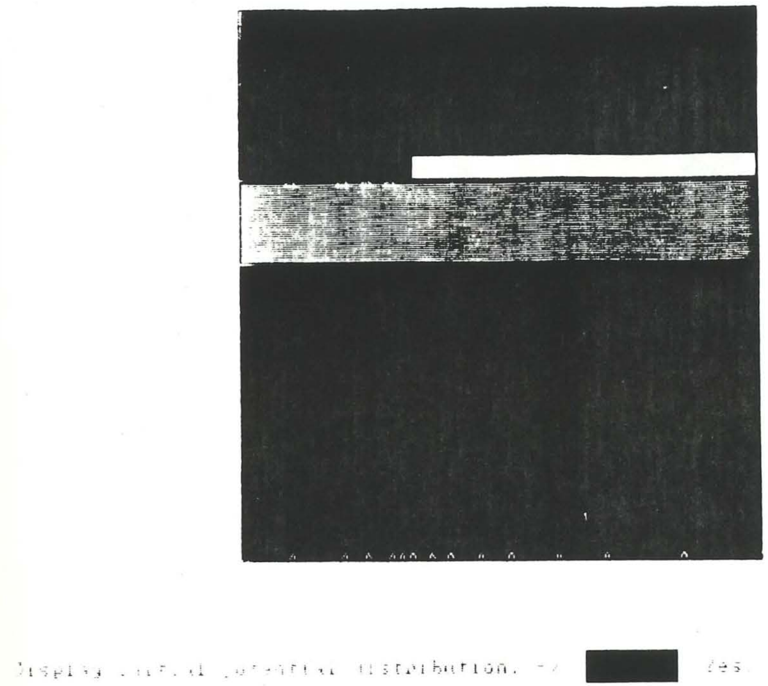


FIGURE 8
"1000 V DMOS TRANSISTOR"

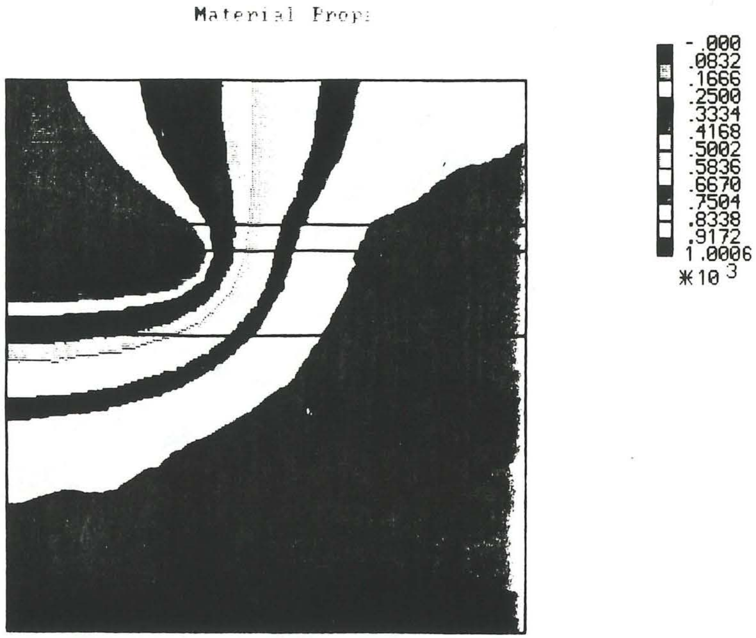


FIGURE 9

POTENTIAL IN 1000 V DMOS TRANSISTOR

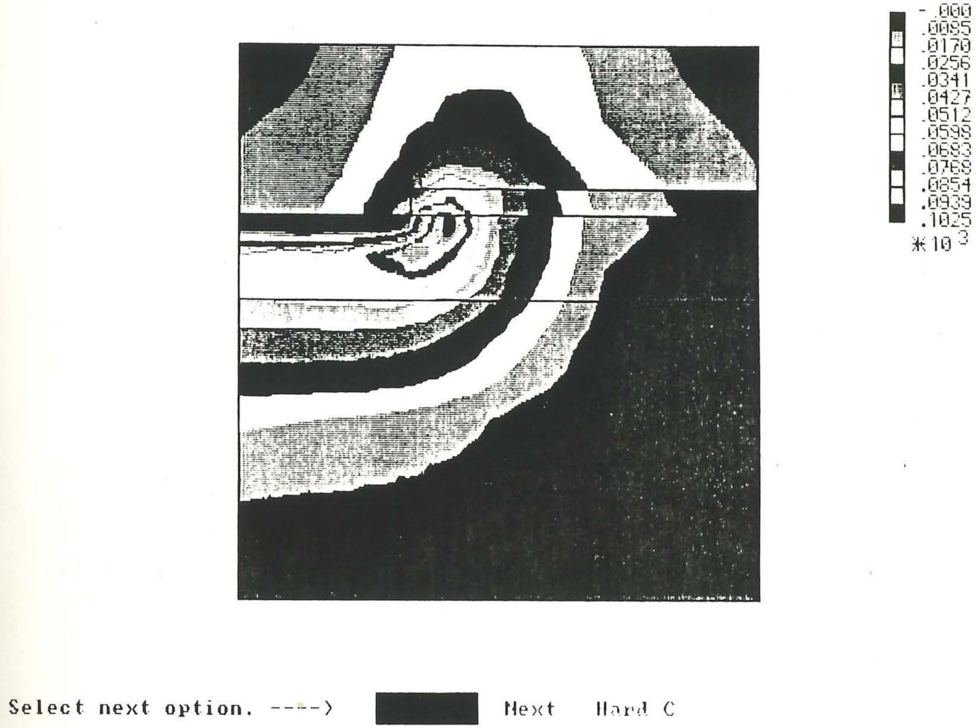


FIGURE 10

"ELECTRIC FIELD IN 1000 V DMOS TRANSISTOR"

In this example only the voltage throughout the device is solved for.

B 1 Micron-channel MOSFET

The second example I wish to show is a 3-Dimensional problem, solved as part of our ESPRIT project. In this case the current as well as the voltage is solved for. This involves the solution of three highly non-linear equations rather than just one.

I would like to finish my talk with a final demonstration, but first, I think something should be said about the future.

7. THE FUTURE

Well, the first thing to be said is that this is in no way going to be an attempt to predict the future. Which particular approach emerges as the best way forward is very often considered an unlikely candidate at an earlier stage. Researchers continue to work away in areas that may appear unpromising, and it must be said that the odds are against any particular one succeeding. I will therefore content myself with some comments on what the indications are and then choose a few topics at random that I think are particularly exciting.

The first thing to be said is that the continued growth in the transistor as the dominant vehicle shows little sign of saturating until the mid to late 1990's. This technology gets cheaper and faster simply be scaling it to smaller size. Every time a more esoteric material or better transistor form has emerged this ability to scale together with the huge and accumulative investment in silicon technology, has beaten off any challenge. To ask the question "Will Gallium Arsenide supersede Silicon?" (it has the potential for five times greater speed and has leap-frogged a lot of technological problems using silicon know-how) is rather like asking whether a titanium alloy will ever replace mild steel. There will be specific areas where cost is less important than performance where other materials will find a place but the mass general-purpose market will be met by silicon for the foreseeable future. After all, any new approach has to cross the great divide between one working effectively to millions working together reliably and with low manufacturing cost.

And now to a few specific areas I think worth mentioning.

1 Low Dimension Structures

The technology of Molecular Beam Epitaxy referred to earlier has enabled new transistor structures to emerge with the potential for very much higher speed. Again using the analogy of materials this could be regarded as a composite and gives rise to the possibility of not 5 times but 200 times speed increase. These materials are strongly anisotropic but have electronic properties in certain directions that are very unusual.

2 Superconductivity

The new high temperature superconductors have a possible application in the quest to increase the performance of Silicon Integrated circuits. When transistors are scaled as we have just described there will come a point when the time it takes to get a signal from one transistor to the next exceeds the time it takes the transistor to switch. When this limit is reached performance will saturate.

However, the delay between transistors will be reduced if the conductors have zero resistance - a state achieved using super-conductivity.

3 Optical Communications

The interconnect problem just referred to may also be alleviated using light-coupling between transistors. If further an optical form of transistor is produced then purely optical computing becomes feasible. Of course communications also then may make use of the interference-free high bandwidth properties of optical fibres.

4 Parallel Processing

My final glimpse of the future is embodied in my final demonstration. The form of transistor I wish to illustrate is called a GTO thyristor. The simulation work was carried out using computer software developed here but in regular use as a design tool at Marconi Electronic Devices at Lincoln. The problem they were faced with is how to design the edges of the wafers so as to minimise the possibility of edge breakdown. Such a design involves choosing an optimum edge profile and filling the space near the edge with silicone rubber which has a higher dielectric strength than air. The effect of an air pocket in the rubber coating is also shown.

CONCLUSION

What is significant about the last demonstration is that it was carried out on a transputer. Parallel computing which the transputer epitomises, is undoubtedly a major and exciting new development. The transistor itself and its development has carried the main burden of these phenomenal developments. Perhaps at the very time when transistor performance is levelling off, advances are being taken on by the computer scientists who in one sense are repaying the debt they owe to the transistor itself.