THE COMPUTER
and the Development
of Science and Learning

**A conference marking the
25th anniversary of
JOHN VON NEUMANN'S
achievement of the
first modern computer**

**Sponsored by
THE INSTITUTE
FOR ADVANCED STUDY
and DAEDALUS**

**6 - 8 June, 1972
Princeton, New Jersey**

Papers and discussion will examine broadly how the electronic computer has influenced the kinds of problems that have been chosen for investigation, and the way they have been defined and solved in selected fields of science and learning in the last quarter century.

The papers, formal comments, and an abstract of the discussion will subsequently be published as an issue of *Daedalus*.

June 6

Afternoon Arrival of Conferees

7:00 p.m. Dinner

8:30 p.m. Introduction
Carl Kaysen
Director, Institute for Advanced Study

Welcoming Remarks
Thomas J. Watson
Chairman of the Executive Committee, International Business Machines Corporation, and member, Board of Trustees, Institute for Advanced Study

Address
von Neumann's Contribution to the Development of the Computer
Herman Goldstine
IBM Fellow

June 7

9:00 a.m.-12:00 Papers

Pure and Applied Mathematics
Stanislaw Ulam
Senior Scientist
Los Alamos Scientific Laboratory
Discussant: Michael Atiyah
Professor of Mathematics
Institute for Advanced Study

Logic and the Foundations of Mathematics
Michael Rabin
Professor of Mathematics
Hebrew University
Discussant: Dana S. Scott
Professor of Philosophy and Mathematics
Princeton University

12:30 p.m. Lunch

2:00-5:00 p.m. Papers

Physics and Astrophysics
K. V. Roberts
Division Head
Culham Laboratory
Discussant: Léon van Hove
CERN

The Applied Physical Sciences
Walter H. Munk
Professor of Geophysics, Institute of Geophysics and Planetary Physics, University of California at San Diego

Jule G. Charney
Alfred P. Sloan Professor of Meteorology
Massachusetts Institute of Technology
Discussant: Gordon MacDonald
Member, Council on Environmental Quality

7:30 p.m. Dinner

June 8

9:00 a.m.-12:00 Papers

Economics
Lawrence R. Klein
Benjamin Franklin Professor of Economics
University of Pennsylvania
Discussant: Herbert Scarf
Professor of Economics
Yale University

The Historical Social Sciences
Charles Tilly
Professor of Sociology and History
University of Michigan
Discussant: David Landes
Professor of History
Harvard University

(*over*)

12:30 p.m.	Lunch
2:00-5:00 p.m.	Papers
	The Biological Sciences Sidney Brenner Senior Member Medical Research Council Laboratory of Molecular Biology, Cambridge *Discussant*: Joshua Lederberg Professor of Genetics Stanford University School of Medicine
	Language, Learning, and Models of the Mind George Miller Professor of Psychology Rockefeller University *Discussant*: Peter Elias Cecil H. Green Professor of Electrical Engineering, Massachusetts Institute of Technology
7:00 p.m.	Dinner
8:30 p.m.	Address
	The Computer and Man's Image of Himself Philip Morrison Professor of Physics Massachusetts Institute of Technology

Computers and the Biological Sciences

Sydney Brenner

Medical Research Council,

Laboratory of Molecular Biology,

Hills Road,

Cambridge.

ENGLAND

Computers have played an important role in some branches of biology but they have not yet come to dominate the field as in some parts of physics. In molecular biology, computers have been extensively used in the determination of the structure of large molecules by X-ray crystallography. This work demands the collection and processing of large amounts of data and, without computers to drive the instruments and to perform the numerical calculations required to derive electron density maps, much of what has been accomplished would have been a technical impossibility. It is easy to extrapolate this experience to other branches of biology; this use of computers is certain to spread if only as part of the continual process of instrumental sophistication that every science undergoes. As more complex problems come to be tackled so the demand is bound to increase for instruments that can record many observations simultaneously and rapidly, store the results and present these in a comprehensible manner. Although scientists place high value on good ideas, ingenuity, and the ability to solve problems on the back of an old envelope, the experimental biologist will learn to use digital computers simply because they have the power to relieve him of the laborious tasks of data collection and processing. Indeed it is quite conceivable that many problems in biology can only be solved by this empirical approach and it is computers that will make this technically feasible. There are now major efforts in applying computer methods to image processing problems, ranging from three-dimensional reconstruction of structure from electronmicrographs to automatic analysis of chromosome patterns. As is well known from the experience with bubble chamber pictures, there are non trivial operations and go far beyond standard numerical computations.

Biologists will have to involve themselves, as the physicists before them, in the difficult problems of handling complex data structures and even in system programming and computer languages.

In all of these examples, the computer is treated as a tool. The hard headed will insist that this is the right way to look at the computer in the laboratory. There is nothing special that distinguishes it from test tubes, microscopes or accelerators, it simply does different things. When the scientist needs it to solve a particular problem he will use it. Many scientists and, in particular, biologists are suspicious of computers, and one must concede that there are enough examples of ridiculous applications to justify their scepticism. With the exception of structure determination, most problems in molecular biology were solved without computers and it is difficult to find examples in which the use of a computer might have hastened or facilitated the solution. Against this rather mundane view of the computer we must place what can only be called the dream of the machine as the universal problem solver, as something more than an instrumental extension of the scientist-programmer. I think it can be agreed that, in practice, this potentiality of the computer has not yet been realized; whether it can and will be realized is another question that I do not want to discuss here. What I want to argue is that biologists have much to learn from computers that goes beyond their use as laboratory tools. To talk at length about applications would be a banal way to commemorate John von Neumann. He was interested in problems that went far beyond the technology of computers. Throughout his work runs the theme of the relation between artificial and natural computing mechanisms and, in particular, he aimed at the construction of general, logical theories of automata applicable to both machines and organisms. His writings are a rich mine of ideas, but their direct influence on biologists has been minimal. Most biologists have never even heard about von Neumann, and very few indeed have actually read his papers. Of course, there has been some influence; the very existence of digital computers has forced new terms into our everyday language and the concepts embodied in words such as programme, control and code are known to biologists and applied by them. I believe, however, that they are used metaphorically and that biologists are still mainly interested in what can be loosely called mechanisms. What I propose to do in the rest of this essay is discuss biological systems from a point of view that is in the von Neumann tradition without making explicit reference to his work.

Biologists deal with very complex systems. They ask three questions about them: How do they work? How are they built and reproduced? How did they get that way? These correspond roughly to the areas of physiology, development and evolution. It is characteristic of modern biology that all answers to such questions are driven down to the molecular level. It is not enough to say that a finger moves because a muscle contracts it; we must also find out how a muscle contracts. We do this by taking muscles apart and making detailed studies of the individual molecular components. From this analysis we then proceed to reconstruct the mechanism. This approach has been successful in providing explanations of many physiological processes. It works because biological systems are structured in a way that allows partitioning. Let us consider one example in more detail. A bacterium synthesizes a large number of specialized small molecules, such as amino acids, vitamins and nucleic acid bases, from a very simple medium containing glucose, ammonia and a few salts. This capacity was once a marvel of living material; today it is simply a bundle of mechanisms. There is a set of elaborate pathways, in which one chemical is transformed into another by specific protein molecules called enzymes, each enzyme responsible for one step. For a long time, some biologists were concerned about the geometrical organization of this system, feeling that it could not function, unless the enzymes were specifically arranged in the cell, and the small molecules were directly handed over, as in an assembly line. We now know that this is not the way it works for most of these processes. There is, in fact, no such thing as a <u>physical</u> address in the cell to which the small molecule has to be directed. It is simply ejected into solution, and, in a cell as small as a bacterium, diffusion will ensure that it will easily find the next enzyme. In the course of its random motion, it will undergo a large number of collisions with the wrong molecules and even with the wrong place on the right molecule. However, these are ineffective, and only when it hits the right place, will there be sufficient binding energy to stabilize the complex and allow the next chemical transformation to take place. This property is embedded in the design of the enzymes, and what has been implemented is, therefore, a <u>logical</u> address and the need for direct physical connections falls away. The methods of biochemical analysis succeed, because, to a very good approximation, the system actually works as sets of partitioned parallel processes. We also know that regulatory connexions between the pathways are implemented in much the same way, by specific small molecules that bind to enzymes and alter their catalytic properties. It is not implemented by all the enzymes consulting some global model of metabolism in the cell.

We can easily see how this method of analysis could be carried out to completion. In principle, it must be possible to do this, since the number of different enzymes in an organism is finite. Suppose, then, we could provide a list of the enzymes and small molecules involved and a specification of their properties and interactions, would we then have an explanation of metabolism? My answer to this question is 'yes', because I do not see any other way to provide one. Notice that this form of explanation does not include any statements about "laws of metabolism", nor is there a "general equation of metabolism"; it contains a description of the structure of the system and a specification of the chemical computations that it performs. It is a set of algorithms specified in the internal language of the system. Only one condition must be obeyed by this algorithmic description and that is a simple, one, subject to empirical test. We must be able to show that any given observed behaviour is computable from the algorithms. Whether or not there exists a different equivalent set of algorithms which could generate the same behaviour is an interesting question but not of immediate relevance to the biologist.

von Neumann was interested in this problem of describing complex automata. He conjectured that, above a certain level of complexity, it may become difficult or even impossible to describe the behaviour of an object and that the object itself was a simpler description. There may be cases "where doing a thing is quicker than describing it, where the circuit is more quickly enumerated than a total description of all its functions in all conceivable conditions." The argument I am making for biological systems is not one of convenience or ease but one of necessity. The prime task of the biologist is to deal with the set of real living organism that are here, now. He is committed to the explanation of a particular class of implementations.

Many may object to this view of biological explanation. There is the persistent belief that since most elaborate systems show some kind of organized or integrative behaviour, any atomistic explanation will be found waiting and that some new principle will have to be invoked to account for these properties. The classic example is that of nervous systems. We observe animals and see that they can do wonderful things, sometimes blindly, sometimes intelligently. I do not accept the view that the physical and logical properties of a nervous system will be insufficient to account for this behaviour and that a mind or consciousness will have to be added. We must treat this case just as we treated metabolism. First, we must proceed to investigate the nervous system directly, determining its structure. This is clearly a system in which a physical address is of importance and a "wiring diagram" is the most valuable piece of empirical information we can try to obtain. Then there is the problem of specifying the set of algorithms to be applied to that structure and that will almost certainly include statements about nerve impulses, synapses and so on. The final task would be to test this algorithmic description against real behaviour. Except for a few elementary examples, this programme has barely begun, and is certain to occupy the energies of many biologists for a long time to come.

Apart from purely technical questions the algorithmic description raises two problems which must be discussed. One is the nature of the internal language to be used and the other involves the methodology of testing it against the real situation. For most physiological processes, the internal language emerges from a study of the physical structure of the system. Biochemical processes are discussed in terms of enzymes and neural processes in terms of nerve cells. Both of these are complex objects in their own right, and there are a large number of questions that can be asked about them. The feature of such objects is that they perform operations at a higher level than collisions of molecules and the breaking and joining of chemical bonds. A heirarchy of such levels exists and our choice of the language is empirically conditioned by what the organism has chosen for itself. It is when we come to the problem of the generation of structures that the choice becomes difficult.

It is now common knowledge that organisms are specified by their genetic material. This takes the form of a large number of messages encoded in the nucleotide sequences of DNA molecules. We can give a fairly detailed account of how these sequences are copied and how they are translated into the amino acid sequences of a protein, and can answer questions such as how is haemoglobin made. But we cannot answer such questions as how is a hand or a nervous system made. The analogy between this biological information processing system and a stored programme digital computer is very striking and it is quite common to find this stated in many biological texts. Often one will find descriptions which are not seriously parodied by the following:

```
PROGRAM MAKEHAND
 CALL MAKEBUD
DO 20 J = 1, 5
20 CALL MAKEFINGER (J)
END
```

Quite clearly, this is not what we want. It is simply a description of what there is in curious prose and is not necessarily written in the internal language of the system. But what is this? The analysis of the internal description that organisms carry of themselves is one of the challenging questions of biology. Organisms are clearly heirarchical structures. They contain organs and the organs are made up of tissues and tissues of cells and so on. One could conceive that they were constructed in this way much as a motor car is assembled from assemblies of subassemblies. A programme for this would have a logical structure that mapped very closely onto the spatial structure of the organism. On the other hand, it is possible to think about it in another way, in terms of procedures rather than objects, such that the mapping is much more complex. The argument here is that organisms have a deep structure which is not easy to deduce from their surface structure but must be studied directly. One way of doing this is to produce alterations in the genetic programmes of organisms and study the consequences of these in great detail. One may hope to discern how the program is partitioned in the organism but it is difficult to predict whether this approach will be successful or not. We are attempting to analyze a programme written in an unknown language for an unknown processor with no comments to help us. However, it should be possible to derive algorithms for devèlopment in the same way as we derived them for function.

How these are to be tested can be relatively easily defined. It is by computation, but clearly the systems considered will be so complex that the use of digital computers will become obligatory. The simulation of such systems will become a standard technique in biology.

For a long time biology remained a purely descriptive science. With the advent of genetics, biochemistry and molecular biology it became analytical and every biologist has learnt to speak the language of chemistry. Our knowledge of living organisms in terms of molecular components and molecular mechanisms is profound but this by itself is incomplete. Organisms, especially complex ones, have a logical structure and biologists will be compelled to learn a new language to deal with this. It is the theory of computation that unites everybody working with elaborate systems and we have much to learn from the experiences of computer scientists.

In one sense, the prospect before us is appalling. It is quite conceivable that the only way to understand organisms is to understand them in detail. They have evolved through a long process of natural selection and their heuristic content may be very high. They may not be unitary and functions and procedures with the same logical content may have been implemented in different ways by different organisms. All of this leads one to suggest that the reason why physicists are astonished by the living world is that they are not engineers. In a deep sense, biology is the study of natural engineering, and biologists cannot seriously search for laws of Nature but only find out what Nature has invented.

'Th Computer and the Development of Science and Learning'

A Conference marking the 25th Anniversary of John von Neumann's achievement of the first modern computer. Sponsored by The Institute for Advanced Study and Daedalus.

Institute for Advanced Study, Princeton, New Jersey, USA. 6-8 June 1972.

* * * * *

COMPUTERS IN PHYSICS AND ASTROPHYSICS

K V Roberts

UKAEA Culham Laboratory, Abingdon, Berkshire, England

To be published in Daedalus

1. INTRODUCTION

It is appropriate at this Conference which is being held to honour the pioneering work of John von Neumann to discuss from a rather fundamental point of view what computers really do and why they should be important to physics and astrophysics: to assess how significant their impact has been in the quarter century which has elapsed since the electronic digital computer was invented and to forecast what may be expected of them in the future. How can one best ensure that computers realize their true scientific potential and that the massive programming and computational tasks which are now being undertaken by the younger generation of physicists are used to maximum effect? Another interesting topic for discussion is the relation between Computational Physics and neighbouring disciplines such as Computer Science and Information Engineering.

It has been said that there are now three main ways of approaching a physical problem: experimental, theoretical and computational. Figure 1 illustrates this point of view which will be examined further in §3. Each of the three approaches has its own characteristic methods of attack, its advantages and limitations, and each extends to physics as a whole as well as to related branches of science. Computational physicists are conscious of this generality and in both journals and conferences they aim to cover the broad field of physics and to bring workers in different areas together. Analogies can be drawn with 19th century Mathematical Physics and these are made more apt by the fact that the Laplace and Poisson equations and Fourier transforms again play an important part.

Several convenient sources are available for those who wish to study the present role of computers in physics and astrophysics, and a brief paper can do no more than to draw attention to these sources and to offer some comments. Leading journals have for several years been the annual review series 'Methods in Computational Physics' [1] and the periodical 'Journal of Computational Physics' [2], both published by Academic Press and edited by a team drawn from the Lawrence Radiation Laboratory at Livermore, California and from UCSD, La Jolla. Major contributions to the development of computational physics have been made by the laboratories at Livermore and Los Alamos, and a review of the Los Alamos work on fluid dynamics has been published by Harlow [3]. An excellent general survey is provided by the recent book 'Computers and their Role in the Physical Sciences' [4], which describes the part played by von Neumann and his colleagues in the origin of the electronic digital computer as well as giving many references to scientific applications in physics and astrophysics. More specialised papers on astrophysics may be found in the journals 'Astronomy and Astrophysics' [5] and 'Astrophysical Journal' [6].

In Europe, a leading role in furthering the growth of computational physics has been played by workers in atomic physics, plasma physics, nuclear technology and high energy physics. The journal 'Computer Physics Communications' [7], edited from Queen's University Belfast and published by North-Holland, was founded in 1969 with the conscious aim of setting up a 'digital scientific literature' , based on a direct analogy with the regular scientific literature [8]. This consists of tested and refereed programs which are distributed on magnetic tape by the associated International Physics Program Library. Standards for documentation and program readability have been proposed [8] and copybook examples published [9]. The Computational Physics Group of the Institute of Physics was founded in the UK at the suggestion of workers from the Culham Laboratory, and held two international conferences in 1969 and 1970 whose Proceedings have been published [10,11]. The success of the 1969 meeting led to the formation by the European Physical Society of an interdivisional Computational Physics Group in which the major part has so far been undertaken by CERN, and the inaugural conference of this group was held at CERN in April 1972 with the title 'The Impact of Computers on Physics [12]. An issue of the CERN Courier [13] which was specially prepared for the conference describes the role of computers in the work of the CERN Laboratory, which can be taken as typical of all the big high energy physics laboratories. The IAEA Centre for Theoretical Physics at Trieste, Italy held a Seminar Course on 'Computers as a Language of Physics' [14] in August 1971 which was attended by many physicists from the developing countries in addition to those from the USA and Europe. Taken together these sources provide a useful 'bird's eye view' of the subject some 25 years after it began.

Figure 2 illustrates the relationship between the three above-mentioned approaches to physics and certain other disciplines. Undoubtedly the deepest form of association, which probably no-one would claim to understand, is that between Theoretical Physics and Mathematics. This is evidenced by the work of Newton and the astronomers and mathematicians of the 18th century, by the mathematical physics of the 19th century, by the connection between geometry and gravitation which is inherent in the theory of general relativity, and most recently by the fundamental significance of group theory and complex variable theory to particle physics which seems to be emerging in our own day. Each discipline stimulates the other and this is also true of the association between physics and Engineering or Technology; theoretical physics contributed Maxwell's equations and nowadays experimental high energy physics would no longer be practicable without the techniques of advanced electrical engineering which are required by particle accelerators. The important stimulus to the development of Information Engineering which was initially provided by the work of von Neumann and his colleagues [4] was subsequently continued by the Los Alamos and Livermore laboratories who encouraged the development of computers such as the LARC and STRETCH and the Control Data 6600, 7600 and STAR. The debt owed by physics to Computer Science is perhaps best illustrated by the extensive use of Fortran which was pioneered by IBM. In the future, computational physics will require and will stimulate the development of advanced new languages, information networks, data storage devices, display systems, pattern recognition techniques and the like. Although the major use of computers lies outside the scientific area, physics can continue to make an important contribution to the progress of computing by its insistence on quantitative measurement and efficiency, abstract symbolism, universality and open publication, and by the scale and difficulty of the problems involved.

The claim implied by Figure 1 is no doubt rather ambitious, and one should ask whether computers are indeed fundamental to science, more so for example than other technologies such as optics or electronics. After all, both the telescope and the microscope led to dramatic advances in our understanding of Nature when they were first introduced and it is doubtful whether computers have yet done as much. The answer does not seem to be entirely clear at the present time. A digital computer can in principle carry out any algorithm that involves only a finite number of logical operations, thus giving considerable added power to the symbols of mathematics. The trouble is that 'finite' to a mathematician does not imply any particular restriction as to size, while to a computer it means $< 10^N$ where at present $N \lesssim 12$ for the fastest machines and will presumably never reach twice this value. In particular, it has recently been pointed out by Emmons [15] that a straightforward finite-difference attack on the problem of 3-dimensional pipe flow at Reynolds number $R_e = 10^7$ can never succeed in practice, since it would take the largest existing computer the full age of the universe to perform a single run. Like Laplace, we are therefore faced with systems of equations for which finite solution algorithms exist but which will forever remain insoluble by direct computational attack. A similar situation did not however prevent the successful introduction of thermodynamics and statistical mechanics, and one may expect that a continuing attempt to use computers to study systems with an infinite or very large number of degrees of freedom will provide an equally fruitful understanding. It is better to travel hopefully than to arrive.

2. AREAS OF COMPUTATIONAL PHYSICS

The purpose of a computer is to process information as we shall discuss further in §4. Now physicists do spend a great deal of time handling information of one kind or another as is illustrated by Table 1, and any impact that computers have on physics must eventually result from this fact.

TABLE 1

A Selection of Information Types and Techniques currently used in Physics

Scientific Literature : printed books and papers
Mathematical symbols, formulae and theorems
Graphs and diagrams
Typescript and manuscript
Pencil and paper, chalk and blackboard
Lectures and conversations
Slides, viewgraphs and films
Experimental measurements
Photographic records
Analogue waveforms
Traces from oscilloscopes and pen recorders
Analogue magnetic tapes
Digital pulses
Electronic logic
Computer printout
Punched cards, paper tape, magnetic tape
Teletypes and visual displays
Computer-generated graphs and movies.

Many of the tools that are currently used for handling scientific information have reached a high degree of sophistication, in particular the mathematical techniques used in theoretical physics, and here it is likely to take a long time before computers can compete on equal terms; one must admit from the outset that the developments in the physical sciences which occurred within 5 years due to the discovery of Schrödinger's equation can hardly be paralleled by those which have occurred within 25 years due to the invention of the electronic digital computer. But in cases where conditions have been more suitable for the introduction of computers, such as the processing of large amounts of digital data from measuring devices or the automatic control of experimental equipment, their impact has been more obvious.

Figure 3 shows the main areas of computational physics, together with certain other areas which might more properly be regarded as belonging to computer science (Languages and Translators), or to information engineering (Hardware and Operating Systems). Again it should be emphasized that the diagram is not restricted to any particular physics field and might indeed be interpreted in a much wider context. For example, many physicists and astrophysicists (as well as other scientists and technologists) are interested in the automatic control of apparatus, and this area covers not only small-scale measuring equipment of many diverse kinds but also the large particle accelerators and remote sensing devices collecting data from the moon or planets. Again, many scientists and mathematicians find it useful to carry out rapid small-scale calculations using 'desk-calculator' facilities at remote consoles. The 'Calculations' area in the diagram has been delineated in somewhat more detail since this is the author's own field.

3. THREE APPROACHES TO PHYSICS PROBLEMS

Let us now briefly discuss the three main approaches indicated in Figure 1.

Theoretical Physics

Theoretical physics and mathematics make considerable use of analogies, many of which are geometrical in character; for example the calculus was originally based on the idea of gradients and areas. Familiar concepts in 2 or 3 dimensions are generalized to n or to an infinite number of dimensions. Both disciplines rely heavily on the use of symbolism, enabling many actual cases to be described by a single algebraic formula. Theoretical physics is position-free since it can survey any portion of space-time with equal ease, for example the inside of a neutron star at some distant epoch.

It is scale-free, ranging at will from the dimension of a quark to that of the whole universe and from 10^{-23} seconds to 10^{10} years, and it is universal in the sense that a single piece of theory such as Coulomb's law or Laplace's equation can be applied to innumerable actual situations.

Theory makes extensive use of linearization. There is almost a motto "When in doubt, linearize". Any linear process is relatively easy to solve by analytic techniques and weakly non-linear processes by perturbation theory. Strongly non-linear processes are much more difficult, whether in hydrodynamics, hadron physics or general relativity. Symmetry and conservation laws are related to one another and play an essential role not only in basic theory but also in practical problems using the method of separation of variables. Complex function theory has a similar dual role; it appears to be fundamental to high-energy physics and to the theory of ordinary differential equations and at the same time is of great practical use in the solution of 2-dimensional problems because of its relation to Laplace's equation. Many of the mathematical 'Methods of Theoretical Physics' have been summarized in the book of Morse and Feshbach [16] and one may envisage that a similar compilation will be undertaken for computational physics in due course.

Approximation techniques are essential. Sometimes this means separating out a few of the many degrees of freedom of a large system, or distinguishing between widely different time-scales as in the method of adiabatic invariants. In other cases such as statistical mechanics the number of degrees of freedom N_F is treated as infinite since this makes the formulae much simpler. Thus theory seems to prefer $N_F \approx 1$ or $N_F \to \infty$.

These are some of the mathematical tools; practical tools include pencil and paper, chalk and blackboard, books, journals and meetings. Theoretical physics is cheap but it requires high intelligence. Another important feature is that theory is self-enhancing; by practising it one becomes a better physicist. This is not necessarily true of experimental physics (which requires the organization of staff and finance, the building of apparatus and the management of contracts), nor of computational physics (which involves struggling with awkward and unreliable computing systems, ill-designed job control languages, much handling of cards and paper and a continual search for errors in the programs). A major task will be to build this feature of 'self-enhancement' into computational physics by improving the techniques and making them more universal and by raising system reliability.

Experimental Physics

Experimental Physics provides the ultimate test and source of information for theory, just as theory provides the equations to be solved by computation. With great ingenuity the scope of experiments and observations has gradually been extended both ways from the human scale to the range 10^{-13}cm - 10^{10} light years in length and from 10^{-23}seconds - 10^{10} years in time. But experimental physics is neither position-free nor scale-free, and the cost of an experiment depends very much on the scale of the phenomenon which is under investigation. Where the expense is high it may be preferable to use theory or computation although experimental modelling is often also of great use. Outside the solar system, and beyond our own epoch, experiments are no longer possible and only observations or computer simulations can be carried out.

No experiment is exact and the potential sources of error must always be carefully examined. This is an accepted part of the experimental ethos which must nowadays be extended to include numerical experiments too; it is perhaps not generally recognized that it should also be applied to the use of approximations in theoretical physics. One advantage of the computer is that it can often be used to check the validity of the approximations that are made, thus providing a second method for testing theory.

Computational Physics

The technique of Numerical Simulation which is used in computational physics combines some of the features of both theory and experiment. Like theoretical physics it is position-free and scale-free, and it can survey phenomena in phase space just as easily as in real space. It is symbolic in the sense that a program, like an algebraic formula, can handle any number of actual calculations, but each individual calculation is more nearly analogous to a single experiment or observation and provides only numerical or graphical results.

To some extent it is possible to solve equations on a computer without understanding them just as one can carry out an exploratory 'botanizing' experiment. This technique is often useful and it seems clear that if the equations of compressible and incompressible flow could have been solved numerically at the time when they were first derived by Euler in the 18th century we should quickly have gained an intuitive understanding of non-linear phenomena such as shocks, vortices, boundary layers, wakes and turbulence. With more complicated processes involving a considerable range of length or time scales it is however essential to make analytic approximations before putting the problem on to the computer, otherwise impossibly large amounts of machine time or storage space may be needed. Not more than about $N_F \approx 10^6$ degrees of freedom can be handled on present-day computers, or $N_F \approx 10^3$ if particles or Fourier components all interact with one another. Thus computational physics can fill in the range between few-particle dynamics and statistical mechanics, in particular by checking the validity of power-series expansions.

Diagnostic measurements on the data obtained from numerical simulation are relatively easy compared to their counterparts in real experiments. This makes it possible to obtain many-particle correlations, for example, which can be checked against theory. Very large quantities of data are however generated by typical 3-dimensional calculations in fluid dynamics, astrophysics, geophysics or plasma physics; for example, a run using 8 variables each of 32 bits, with a space mesh of size (100x100x100) and a duration of 1000 timesteps, will produce 2.5×10^{11} bits of data. Similar storage requirements occur for experimental data in many fields of physics and on-line devices holding up to 10^{12} bits are currently available [13]. Holographic methods for rapid access to much larger volumes of data are envisaged in due course. With numerical simulation it may often be more economic simply to repeat the run, but irreplaceable data from geophysical or space physics observations clearly must be stored. An added incentive to the construction of large-scale storage devices will come from the attempt to make the scientific literature available on-line.

Numerical simulation is particularly suitable for non-linear, non-symmetrical phenomena where the usual techniques of Fourier transformation and separation of variables do not apply (as in weather calculations), but often the programs are easier to write and the calculations go much faster in simple situations such as rectangular Cartesian geometry with periodic boundary conditions. Computers are also good at handling situations where considerable accuracy is required, as in the optimization of nuclear reactors, and situations of great complexity described by many coupled equations.

Synergetics

Sometimes one method of approach will be more appropriate and sometimes another; frequently they will work in pairs and at times all three methods must be used together. This mutual support is termed Synergetics and has been discussed by Zabusky [17] in connection with the Fermi-Pasta-Ulam problem. An example where computational techniques are particularly appropriate is in the solution of the equations which describe the internal structure and evolution of stars [18,19]. These equations are complicated and non-linear but they are well-defined, and provided that spherical symmetry can be assured they are well within the range that computers are able to handle. On the other hand analytic methods have difficulty because of the non-linearity, whilst it is clearly awkward to do experiments or even to make observations (except with neutrinos) in the interior of a star. Detailed calculations of internal stellar structure and evolution have therefore been carried out and compared with observations on mass, luminosity, spectral type, pulsation period, light curve and other data, all of which refer only to the outside of the star. It is however noteworthy that in the one case where internal observations can be made, namely on the neutrino emission from the sun, there does appear to be serious disagreement and the onus is now on theory to explain this [20]. A prime goal of both experiment and computation in handling complex many-body phenomena is to find striking new regularities or paradoxes for theory to explain. The book by Betchov and Criminale on 'The Stability of Parallel Flows' [21] gives a good account of the way in which analytic and computational techniques can support one another in one branch of fluid dynamics.

4. CLASSES OF INFORMATION

It is characteristic of the electronic computer that it deals primarily with digital information, although capable of handling many other types of information as indicated in Figure 4. All forms of digital information are freely convertible into one another but in many other cases the transformation is essentially one-way, being much easier out of the digital form than into it. Thus it is relatively straightforward to make a computer talk but much harder for it to understand the human voice. It can draw graphs but not so readily interpret photography. It can print books but not easily read them. Therefore there is a natural tendency to replace each type of information by its digital equivalent so that it can readily be transmitted, duplicated, manipulated and displayed, all other forms of information being then regarded as evanescant versions intended only for immediate display to humans. Analogue-to-digital conversion can be carried out readily in either direction with some loss of accuracy, so that measuring devices equipped with A/D convertors will increasingly be directly connected to the computer, and wire chamber data may eventually replace bubble-chamber photographs [13].

Table 1 will thus gradually be unified although it is not clear how near to completion this process will go.

If this development has not proceeded very far in physics as yet, it is partly because of the vast amount of data with which the average physicist habitually has to deal. A typical physics text in his private library extends to 500 pages, with some 2500 characters on each page, that is about 10^7 bits altogether. His complete private store of information is likely to be of order 10^{10} bits. This may be compared with the private allocation of only about 5×10^6 bits of direct-access storage which is presently available to the author at Culham. It is clear that taking into account the requirements of the central library and other public information, even a moderate-sized laboratory will require several x 10^{12} bits of direct access storage before computers become fully effective. It is true that infrequently-used information can be archived on to magnetic tapes but in practice this causes great inconvenience. Another obvious point is that books are pleasant to handle; they employ lower-case characters, mathematical formulae, diagrams and photographs, and they can be taken home or read on trains. These advantages will be less apparent when we have cheap portable graphic displays that can be attached to any telephone.

Mathematics and physics have a peculiar relationship to graphics. Many of the basic ideas of mathematics are geometrical in origin and physicists use graphs in innumerable different ways. Computers can easily draw diagrams using either pen recorders or cathode-ray tubes, although there has been a tendency in the past to rely mainly on numerical output from line printers which are actually more expensive than many graphical output devices. It is now possible to buy cheap, silent remote terminals capable of displaying up to 300 vectors or 1200 characters/second, which can be linked to a computer or telephone line and can also be provided with a local hardcopy facility. Having such a device in his own office is likely to cause a major change in the way in which the physicist uses a computer, since not only will it be possible to generate dynamically-changing diagrams but also to extend the power of programming languages by using (software-generated) Greek symbols and other mathematical signs which have so far been missing. It will be interesting to see how mathematicians attempt to display the properties of n-dimensional space now that they have this new tool. (3 dimensions can readily be achieved using polaroids, of course).

Graphical input is a fundamental problem, and the analysis of bubble-chamber photographs produced at CERN is responsible for much of the scientific computing that is currently undertaken in Europe. It has led to the installation of the fastest available computers (IBM 360/195 and CDC 7600), and to a great deal of work on pattern recognition.

Another area of interest to physicists is high-speed data transmission, which should occur at approximately TV rates so that a worker at a distant laboratory or university can view an on-line computer movie being generated at a central facility. Zacharov reported recently [22] that he has set up a 10^7 bits/second microwave link at the Daresbury laboratory in the UK, and tested it by bouncing the OS/360 operating system backwards and forwards several time before using it to process jobs successfully. Pasta [23] outlined the US concept of a National Science Computer Network which could supply a complete information service to a university using a single commercially available line of about 50 Kbaud capacity.

Although the scientific literature will no doubt be made available on-line in due course, this policy has many dangers which should be carefully examined. The present reliability and indestructibility of the literature stem partly from the fact that many copies of each textbook and journal are distributed; these are freely intelligible to all scientists, are open to public criticism and are virtually impossible to alter or destroy. A computer data bank is much more vulnerable, and it is hard to see how the scientific works of antiquity would have survived if they had been held in digital form in one central location. The destruction of the library at Alexandria was a serious loss to science.

5. COMPUTER SIMULATION OF CLASSICAL SYSTEMS

Computers are extensively used to simulate the time-dependent behaviour of physical systems described by classical laws. Figure 3 gives a partial breakdown of this area of activity: two fundamental areas of interest are evidently the solution of sets of coupled partial differential equations governing the behaviour of fields or continuous fluid media, and direct particle-particle interactions as between stars in a globular cluster or molecules in a fluid. Since it is impracticable to follow the N(N-1)/2 mutual interactions of more than about $N = 10^3$ particles on present-day computers, an important hybrid technique has been evolved which uses both particles and fields. Typically one follows the motion of up to 10^6 point charges or stars obeying Newton's equations

$$\frac{d\underline{x}_i}{dt} = \underline{v}_i \qquad (1)$$

$$m_i \frac{d\underline{v}_i}{dt} = e\,\underline{E}\,(\underline{x}_i) \qquad (2)$$

where (say)

$$\underline{E} = -\operatorname{grad}\varphi \qquad (3)$$

is the electric field computed by differencing the potential φ on a discrete space mesh. The potential is in turn obtained by solving Poisson's equation

$$\nabla^2 \varphi = -4\pi\rho \qquad (4)$$

the charge density ρ being calculated from the positions of the charges themselves. This method distorts the close encounters between individual particles, but fortunately these are not too important in collective phenomena which are governed by long-range interactions of the Coulomb type. It is less suitable for studying short-range encounters between the atoms or molecules in the solid or liquid state, but here the hybrid method is not really needed because almost all the interactions are sufficiently distant to be neglected and only the forces between near neighbours need to be evaluated. Another significant technique is the Monte Carlo method which is used for solving complex problems in neutronics or Knudsen flow. This is based on the idea that the results of individual collisions are only statistically predictable and can therefore be computed with the aid of a random-number generator.

Since much of the universe is governed by classical, non-linear physical laws there is clearly a rich field for the application of computer simulation and Table 2 lists a few of the topics which have either been investigated or

TABLE 2

Some Topics in Classical Numerical Simulation	
Earth's dynamo problem	Aerodynamics
Mantle convection	Structure of real shocks
Seismology	Fluid dynamics
Oceanography, tides	Turbulence
Meteorology	Thermal convection
Ionosphere and Space physics	Plasma physics
Solar flares, sunspots	Magnetohydrodynamics
Stellar atmosphere	Electron tubes
Stellar structure and evolution	Particle accelerators
Pulsars	Industrial flow processes
Stellar clusters	Nuclear reactors
Cosmic gas clouds	Nuclear weapons design
Galactic structure; spiral arms	Nuclear weapons effects
General relativity	Statistical mechanics

are being considered at the present time. Another useful classification is indicated in Table 3.

TABLE 3. Types of Computer Simulation

I. Idealized theoretical studies
II. Investigation of real physical processes
III. Design of apparatus or equipment

It is appropriate at this point to repeat the caution already given in §1. Most of these classical systems possess a virtually infinite number of degrees of freedom N_F, with space and time scales varying by factors of 10^9 or more. It is by no means obvious how to compress the physics into the $N_F \approx 10^3$ to 10^6 degrees of freedom that can be handled by the computer, and a rather deep theoretical understanding is needed in order to devise suitable finite methods and to investigate the numerical errors which occur. There is an interesting analogy, which deserves further examination, between linearized modes on a finite difference mesh and waves on a periodic lattice in solid state physics. A cubic mesh with spacing Δs cannot represent wave-numbers with components having absolute values greater than

$$K_{max} = \pi/\Delta s \tag{5}$$

so that only a finite region R_o of $\underline{K}$-space near the origin can be described. If two modes $\underline{K}_1$ and $\underline{K}_2$ couple non-linearly to give a third mode

$$\underline{K} = \underline{K}_1 + \underline{K}_2 \tag{6}$$

which lies outside R_o, then its energy must be 'reflected' back inside this region by a type of Umklapp process which is known as 'aliasing'. This is particularly awkward in 3-dimensional turbulence studies where energy should cascade towards the shorter wavelengths [24-26].

A different and more fundamental problem [27] occurs in 2-dimensional incompressible fluid dynamics where energy tends to move the other way, from the shorter to the longer wavelengths [24-26]. Since the earth's atmosphere behaves in many respects like a 2-dimensional fluid there is currently some reason for believing that weak small-scale disturbances may within a finite period build up to produce macroscopic effects, making the weather <u>unpredictable in principle</u> after a time which appears to be of order 14 days. This hypothesis can readily be tested on the computer by introducing a small disturbance at some point within a previous calculation and then repeating the run to see what difference it makes.

In partial contradiction to this idea are many numerical experiments in

2 dimensions in which strikingly regular phenomena are seen to arise as the final state of unstable non-linear flows [28]. Figure 5 shows the results of some calculations carried out by Christiansen and the author [29], in which the motion of clouds of point vortices was followed using the particle-and-field method. In many cases the final state consists of one or more stable interacting vortices of finite area whose properties ought to be predictable by simple theoretical arguments, the only uncertainty being a phase degeneracy due to the periodic symmetry in cases (a) and (b) and rotational symmetry in case (c). Sometimes an apparently stable state lasts for only a finite time, after which it breaks up as shown in case (c), and subsequently forms a new pattern on a larger scale. Similar processes of 'vortex condensation' are observed experimentally and this is currently a fruitful field for synergetic study by all the three approaches indicated in Figure 1.

Phase changes, for example those between the solid and liquid states, can be studied on the computer either by the method of Molecular Dynamics which follows the motion of a few hundred interacting particles [30] or by a modified Monte Carlo method which counts configurations [31]. This is a non-linear situation in which it is difficult for theory to make progress and the calculations by Alder, Wood and others have been instrumental in casting doubt on some of the current theories and in supporting earlier ideas such as those of Van der Waals. The fact that only a small number of particles can be used encourages a careful rethinking of some of the basic concepts of statistical mechanics.

An important early success was the work of Fermi, Pasta and Ulam [32] on the dynamics of non-linear or anharmonic lattices of point particles. A recent paper by Zabusky [17] points out that this has blossomed into three distinct but overlapping areas :

1. Thermal conductivity in non-linear crystals.
2. Mixing, ergodic or 'stochastic' behaviour and the approach to thermodynamic equilibrium.
3. Non-linear wave propagation, energy focussing and sharing, solitons and the Kurteweg-de Vries equation.

Again it is found that the behaviour of non-linear systems is much richer than might have been expected and that previously unpredicted regularities occur which then stimulate fundamental theoretical work such as that of the Princeton group [33]. Computer movies or animated on-line interactive displays can play an important role in bringing these regularities to light.

In the technological field two of the most important and successful applications areas have been the design and optimization of atomic and thermonuclear weapons and of fission reactors, and for nearly 30 years this work has pushed forward the development both of new computer hardware and also of many basic numerical methods [3,34]. It is interesting to examine the reason why these particular calculations should have been so successful, and in the case of weapons hydrodynamics it appears to be that the flow is constrained by technological requirements to remain simple, in contrast for example to the complex flow past the body of an aircraft. It could therefore be handled adequately even by the earliest computers provided that arrangements were made to deal with shocks [34,35] and fluid interfaces [3] , while aircraft and ship design still relies heavily on the use of analogue devices such as wind tunnels and model tanks.

In astrophysics it is again the simple 1-dimensional calculations on stellar structure and evolution [18] and the dynamics of variable stars [36] that have been most successful, although much important work has been done on the dynamics of stellar clusters [37] and on the origin of the spiral arms of galaxies [38]. Stellar clusters are interesting from the statistical mechanics point of view because two point masses can release an unlimited amount of energy by moving sufficiently close together; the ultra-violet catastrophe which was removed for charged elementary particles by the introduction of quantum theory still exists in gravitational theory and can be seen in action in computer runs. The problem of the stability of the spiral arms still remains after a considerable amount of theoretical [39] and computational work [38], and other interesting areas for numerical simulation in astrophysics include solar flares, rotating stars, X-ray stars and pulsars [40], and unsymmetrical initial-value problems in general relativity [41]. In all these areas the application of computers is favoured by the difficult non-linear 3-dimensional nature of the problems and by the impracticability of real experiments and sometimes even of scaled models. It is now feasible and economic to perform 3-dimensional magnetohydrodynamics and astrophysics calculations on a (100x100x100) space mesh using a computer such as the Control Data STAR-100, but further work will be needed on programming techniques, numerical analysis and display methods before a marked impact on the progress of astrophysics begins to be felt.

6. TWO USEFUL ANALOGIES

Because computational physics is a comparatively young branch of science it is useful to draw upon analogies with older, better understood disciplines to see how it can best be developed. One interesting analogy is that between a numerical experiment and a real laboratory experiment. Surprisingly enough, their costs and the time and effort needed to set them up are often quite comparable, and there is every reason to be as efficient and sophisticated as possible in large-scale computational work.

The computer program corresponds to the apparatus of the experiment, and it should be equally well designed, constructed and tested according to sound engineering principles. Just as with a major laboratory apparatus it will often be necessary to employ substantial design and programming teams and to use techniques such as modular construction and critical path planning. A program can be equipped with built-in instrumentation for test purposes, and it should be well documented so that it can readily be understood by future workers after a period of several years, or modified for use on other computer systems or other physics problems. Careful checks should be made on numerical accuracy, the empirical measurements being compared with theoretical estimates.

A straightforward simulation run which generates nothing but vast quantities of numbers is analogous to a laboratory experiment in which no measurements are actually made. To be useful, the raw apparatus must in each case be supported by adequate diagnostics. An advantage of the numerical experiment is that detailed measurements can be made without any disturbance to the physics at all, although often at some cost in computer time. Properties such as many-particle correlation functions can be examined and compared against theory which it would be impossible to observe in any other way. Alder has pointed out [30] that in such a case one also knows the exact law of force, which would not be true of many real experiments, so that computing can provide a more accurate check of the approximations made in the theory.

Modern experiments frequently use modular, pre-packaged diagnostic equipment with standard interfaces, e.g. the CAMAC series. Corresponding developments are taking place in numerical simulation in order to reduce programming delays. It is also practicable with a multi-programming computer system to carry out simulation runs 'on-line', monitoring them from a control desk equipped with one or more keyboards and visual displays. A run can be suspended to allow time for thought, it can be restarted from a previous point in order to test alternative physical assumptions, or it can call upon the physicist himself to make decisions which are too difficult to program. It seems likely that some of the more complex astrophysical situations may best be simulated in this way.

A second useful analogy is between a computer program and part of the mathematical literature, e.g. a textbook. Journal articles and books dealing with complex subjects are carefully designed to be as intelligible as possible, and many of the same techniques can be taken over in the structure and documentation of physics programs [8]. So far as practicable all workers in the field should be able to share a common pool of notation, conventions, programs and subroutines just as they share the techniques of mathematical physics, and it is for this reason that the journal 'Computer Physics Communications' and the International Physics Program Library [7] have been set up. Physicists and mathematicians are beginning to recognize the need for compendia of tested and published subroutines, analogous to the comprehensive 'Handbook of Mathematical Functions' of Abramowitz and Stegun [42].

7. QUANTUM MECHANICAL CALCULATIONS

Extensive quantum-mechanical calculations are carried out in atomic, molecular and solid-state physics, in quantum chemistry and in nuclear and particle physics. Although these clearly cannot be discussed adequately in a single paper it may be illuminating to make some comparisons with classical computational physics.

For practical purposes the mathematical laws governing the behaviour of electrons and nuclei in matter can be considered exactly known, being represented by the Schrödinger or Dirac equations together with perturbation theory in quantum electrodynamics and a small correction where necessary for the finite charge distribution on the nucleus. If these equations could be solved numerically with sufficient precision then vast areas such as atomic and molecular spectroscopy, electronic and atomic collision phenomena, quantum chemistry and solid state physics could simply be handed over to the computer. Our earlier considerations might suggest that this could never be possible because of the large number 3n of dimensions and therefore of 'degrees of freedom' associated with Schrödinger's equation

$$-i\hbar \frac{\partial}{\partial t}\psi(\underline{r},t) = \left\{\sum_i \frac{\hbar^2}{2m_i}\nabla_i^2 - V(\underline{r})\right\}\psi(\underline{r},t) \qquad (7)$$

for an n-particle system. In fact, however, much more precise results can be obtained in quantum theory than would be the case for analogous partial differential equations in classical physics, and it is instructive to ask why this should be.

The first point to notice is that the explicit time-dependence represented by equation (7) is rarely needed; we require mainly eigenvalues and eigenfunctions of stationary states together with transition probabilities between them. The real time t can therefore be formally replaced by an imaginary time τ so that (7) can be written symbolically as

$$\frac{\partial \psi}{\partial \tau} = (D\nabla^2 - V)\psi\,, \tag{8}$$

analogous to the diffusion equation governing the buildup of the neutron population in a supercritical assembly. A region of negative potential V corresponds to a 'source' region where fission dominates absorption, while positive V corresponds to a neutron-absorbing region. After a long 'time' τ the fastest-growing mode

$$\psi \sim e^{E_o \tau} \tag{9}$$

dominates, where $-E_o$ is the energy of the ground state, so that a straightforward initial-value calculation in imaginary time provides both the ground state energy level and wave function.

This idea can be traced back a long way, at least to Feynman's thesis (1942). It has recently been used as the basis for a number of prototype computations using mesh techniques [43], Monte Carlo [44] and Feynman path integrals [45]. One of the problems in using Feynman paths to solve (7) directly is that their amplitudes are complex and interfere destructively so that a numerical sum over paths is unlikely to converge. With equation (8) the amplitudes are real and most of the difficulties no longer arise. The main conclusion we wish to draw here however is that because (8) is a diffusion equation the ∇^2 term ensures that the high wave-number modes rapidly disappear, so that quantum-mechanical functions tend to be smoother and thus more easily calculated than their counterparts in classical physics where for example in turbulence the amplitudes of short wavelength modes continually increase. An alternative and more usual interpretation is that plane-wave components of high momentum $\underline{p} = \hbar \underline{k}$ in the solution ψ are inhibited by energy conservation.

Another noteworthy fact is the remarkable stability, precision and reproducibility of quantum-mechanical structures such as atoms or molecules compared to anything that exists in classical physics. Although we talk about the 'uncertainty principle', the future state of a complex quantum-mechanical molecule is actually much more ordered and predictable than any structure of interacting classical particles governed by Earnshaw's theorem could ever be and this is why such a molecule is able to serve as a reliable carrier of genetic information. This point was emphasized by Schrödinger in Chapter IV of his book 'What is Life' [46], published in 1945. For our purposes it suggests that detailed numerical computations of complex atomic structures are meaningful and have some chance of success.

The actual solution of equation (7) is an excellent example of the synergetic approach. It is doubtful whether any mathematician faced with the many-particle Schrödinger's equation could have predicted the existence of complex atoms and molecules ab initio, and certainly not by using a computer. The development was in fact guided by a great deal of chemical, spectroscopic and X-ray evidence from experiments, together with the particle concepts of classical physics and the older quantum theory. This included the idea of closed electron shells and the explanation of the periodic table given by Bohr in 1922 [47], while Hartree [48] in 1923 and Lindsay in 1924 [49] had already begun to introduce the concept of a self-consistent field [49]. By the time that Schrödinger's equation was put forward in 1926, therefore, a considerable amount of information was available concerning the qualitative nature of its solutions, and it was natural to introduce techniques such as the Rayleigh-Ritz variational method, the 1-electron model, the Hartree and Hartree-Fock methods and the Born-Oppenheimer approximation in order to make numerical or analytic calculations practicable.

An important feature of these solutions is their approximate separability. Complex systems can be broken down into weakly interacting subsystems, and solutions of the multi-dimensional equation (7) can be represented by anti-symmetrized product wave-functions of the Hartree-Fock type with some allowance for configuration interaction or correlation. As a consequence the amount of information required in a numerical solution is enormously reduced, so that for sufficiently large molecules it presumably depends only linearly on the number of particles n instead of exponentially. The appropriate breakdown into subsystems depends in the first place on a combination of experiment, theory and intuition but it can increasingly be tested quantitatively on the computer.

These mathematical and physical properties of the solutions have the practical consequence that 'experiments' or 'measurements' in atomic physics or chemistry which are made on the computer are gradually becoming more and more competitive with those made in the laboratory. Naturally this development has gone furthest where the calculations are relatively straightforward and laboratory experiments are particularly difficult to carry out, and astrophysics has relied for some time on atomic data obtained from numerical calculations. Because of the considerable amount of programming and computing required, as well as the possibility of a breakdown into physical sub-systems which can be combined together in many ways, this field of physics depends for its continued growth on good organization and international collaboration. Easy-to-use, general-purpose programs and subroutines should be available in published form, together with large data banks storing integrals, wave-functions, and sub-system properties in order to avoid unnecessary duplication. According to Burke [51] :

> "Firstly, in those areas where the theory is now on a fairly firm foundation we can expect the development of very good atomic codes which will allow astrophysicists, plasma physicists etc. to obtain answers to their atomic questions on request. Already a number of such codes are close to completion and the best way of releasing them to the physics community is under active consideration. It is clear that the preparation of input data and the analysis of output from such codes present formidable logistics problems and some thought has to be given to the design of suitable control programs. Recent efforts in this direction have been made using an IBM 2250 graphical unit to display tabular or graphical material stored in a data bank and to initiate runs of the appropriate computer code if the required data is not already available [52]. The second area which is under active development at the moment is the design of an on-line retrieval system which will use as its data base suitably verified data extracted from the literature [53]. One of the main problems here is to study the best way of compressing and presenting the vast amount of information on quantities like potential energy curves, cross-sections, oscillator strengths etc. which are at present or will shortly become available. Both of these developments are seen to be aimed mainly at the physicist in some other field who needs a number to enable him to interpret data of interest to him".

Wahl [54] has given a good account for the non-specialist, with many references, of the use of computers in quantum chemistry. He points out the great advantage of being able to study collision processes in graphical form:

> "We can 'freeze' our numerical model and 'look' at any stage of a process. For example we can now watch a molecule form or atoms collide in terms of their changing electronic charge density continuously being displayed on a cathode ray tube controlled by

digital computers during the chemical process numerically under way".

Although such calculations are expensive, films produced in this way [55] may be expected to have great educational value in demonstrating the qualitative and quantitative meaning of many of the models and approximations that are used, and this applies generally throughout much of physics.

In nuclear and particle physics the use of the computer is somewhat different because here the basic laws are less precisely known. The 'equipment' required is however much the same, and a noteworthy development is the DELPHI-SPEAKEASY system set up by Cohen [56] at the Argonne National Laboratory which was originally intended to serve the field of nuclear shell-model studies and has since become a generalized program for theoretical physics.

As a final comment in this discussion of quantum mechanical calculations we recall the highly and increasingly accurate work on the 2-electron bound state problem which has been carried out since the task was begun by Hylleraas[57] in 1929, culminating in the most recent calculations of Pekeris and colleagues [58] published last year which yield twelve places of accuracy for the ground state of the helium atom. One of the aims here is to check the higher-order predictions of quantum electrodynamics.

8. DATA PROCESSING AND ON-LINE CONTROL OF APPARATUS

Electronic processing of experimental data or control information becomes worthwhile and even essential when the parameters lie outside the human range. This may occur for many reasons including the vast amount of repetitive data processing required, the speed of accuracy of response that is needed, and the inaccessibility of the equipment. No one field of physics can be singled out but amongst those at the forefront in this application of computers may be mentioned:

Space physics
High energy physics
Radio astronomy
X-ray crystallography
Seismology.

A standard pattern of computer network for physics laboratories is beginning to establish itself. This typically includes a powerful central computer C coupled to a smaller 'front-end machine' F, with access to a large data-bank D.

Each sizeable experimental apparatus or piece of equipment has one or more data-processing and control computers of its own, and these are linked via F to both C and D. Also attached to the front-end machine F are teletypes and displays throughout the laboratory, together with graphical hardcopy devices and possibly a microfilm recorder in addition to routine peripherals such as card readers, printers and magnetic tape decks. Remote job entry equipment may be situated at points some distance from the main computer building, and F will usually be connected to the automatic dialled telephone network as well as possibly to other computer systems via high-speed data links.

The actual situation is very fluid at the present time because of the large quantity of incompatible hardware and software that already exists and is gradually being coupled together. Standardization is however essential for further progress and will therefore presumably take place. Another important requirement is reliability. Large computer systems are known to break down every few hours and to require much operator intervention, and this clearly must not be allowed to affect the working of an entire laboratory. The individual elements of the network must therefore be designed to work separately so far as possible, with some duplication of essential services. An attractive possibility is to delegate some of the functions previously carried out by C such as file editing, compilation and small interactive calculations to a smaller computer which can function reliably and unattended on a 7-day, 24-hour basis. The design and operation of a laboratory or even a national or international network [23] is increasingly a problem in Information Engineering.

There are obvious parallels between particle physics and atomic spectroscopy on the one hand, and between radio and optical astronomy on the other. It is interesting to enquire why in each of the two new disciplines it is necessary to rely on extensive data processing where previously in the older disciplines a simple photographic plate sufficed, a development which is perhaps related to the increase in both the scale and cost of the equipment.

In each case it appears to be wavelength of the particles or quanta that is the important parameter. In particle physics quantum events have to be analysed individually in great detail, and the energy analysis must be carried out by large-scale magnetic fields rather than by diffraction gratings, interference filters or prisms which are only available at much longer wavelengths comparable with the atomic dimensions of ordinary matter. To minimize $\sqrt{N}$ statistical fluctuations it is evident that millions of events are required which only a high-speed computing system can handle. Conversely, the very long wavelengths used in radio-astronomy also require large-scale equipment in order to get adequate angular resolving power, culminating in the use of base-lines thousands of miles long, the data being recorded on magnetic tape and synchronized by atomic clocks. Here the information consists of continuous classical waveforms rather than statistical quantum-mechanical data, but it is on such a large scale that only a computer can analyse it.

Once data has been accumulated and processed it must be stored in convenient form for future interpretation. There are exciting possibilities here, since previously-recorded experimental information can be summoned up from a console, analysed according to some newly-designed algorithm, and the results displayed on-line in graphical form. A good example was demonstrated recently in the BBC TV program 'The Restless Earth' in which seismic data that had previously been recorded by the large detection array in Montana was being analysed on-line at MIT, the phases being adjusted at will to 'point the telescope' in any desired direction and so locate the focus of an earthquake. Problems arise from the enormous quantity of experimental and observational data that is potentially available, and the need for adequate indexing schemes and new languages to enable it to be located and analysed so that it can be exploited effectively by physicists. The CERN magnetic tape library is currently increasing at the rate of 10,000 tapes a year, while NASA is said to be receiving 100,000 tapes a year of data from satellites in orbit and now to have about 750,000 tapes altogether in its library [13]. Handling such large quantities of information places a considerable burden on the computer operating staff, and forms of analysis which require the use of a small quantity of data from each of many records will hardly be practicable until truly random-access on-line data storage is in use.

Computer processing of televised pictures from the moon and Mars is now quite familiar and it seems clear that for those wavelengths that cannot propagate through the earth's atmosphere the immediate future of observational astronomy lies with automated, orbiting observatories which telemeter their data back to stations on the ground and can where necessary be controlled on-line by earth-based observers. Automatic acquisition and processing of data from ordinary optical telescopes are also being introduced.

9. COMPUTER-ASSISTED DESIGN OF APPARATUS

Many types of physics apparatus are now being designed with the aid of computers. On-line 'desk-calculator' facilities are very popular using ordinary teletypes or visual displays connected to a central computer, although these meet strong and healthy competition from individual desk-calculators which are often more convenient to use and can be equipped with stored programs, function keys and plotting devices. The use of interactive graphics is likely to prove increasingly important in apparatus design as well as in many other areas of computational physics, but seems to have been hampered in the past by the high cost of graphics terminals as well as by the undue complexity of the software. A storage tube such as the Tektronix 4010 is now quite inexpensive and on some computer systems can receive its character or vector input directly from user-written Fortran routines in A-format, so that a substantial increase in this kind of application may be expected.

Monte Carlo methods are widely employed for the optimization of particle detection apparatus used in high energy physics, and another interesting application is in the design of high-vacuum equipment where individual molecules follow straight-line paths between their collisions with the walls.

10. ALGEBRAIC MANIPULATION

Mathematics provides many standard tools for the physicist [16]. Some of these are purely mechanical such as the rules for differentiating elementary functions, others are somewhat more heuristic like the rules for elementary or contour integrals; many are very intuitive indeed. But all these tools are used over and over again throughout the whole of theoretical physics and so there is good reason for trying to mechanise them in order to remove unnecessary labour.

Much work has been done by computer scientists in the difficult field of Machine Intelligence, but little of this has produced new results in mathematics or physics as yet. Physicists themselves have concentrated on areas where relatively straightforward algebraic manipulations must be carried out a large number of times. These include the evaluation of products of γ-matrices or Feynman diagrams in 6th or 8th order quantum electrodynamics [59], the handling of tensors in general relativity, the evaluation of coupling coefficients in atomic calculations [51], and the automatic generation of computer programs to solve coupled sets of partial differential equations [60].

These successful applications have emphasized some of the difficulties including the lack of an entirely suitable programming language, the tendency for intermediate expressions to become very long and so consume much storage space, and the difficulty which the computer finds in performing some operations which are quite obvious to a man such as

$$\begin{aligned} \cos^2\theta + \sin^2\theta &\rightarrow 1 \\ a^2 + 2\,ab + b^2 &\rightarrow (a+b)^2 \quad . \end{aligned} \tag{10}$$

A temporary solution would seem to be the development of a 'power-assisted algebra' facility, in which the computer would do the actual hard work but many of the difficult decisions or suggestions would be made by the physicist user. This is impracticable with batch processing systems, and not too easy with teletypes because of their noise, low speed and restricted character set, but it is becoming increasingly feasible with fast visual displays. One can readily imagine displays which would be equipped with hardware-generated Greek symbols and standard mathematical signs, together with keys for frequently-used operations, but much of this can already be simulated by software. A light pen is an excellent tool for specifying the path of a contour integration.

11. COMPUTERS AND THEORETICAL PHYSICS

Finally let us discuss a paradox. Theoretical physicists and mathematicians have always played a leading role in the development of computers, from Leibnitz and Babbage to Turing and von Neumann, and the actual computer installations themselves have traditionally been run by the mathematics departments of universities and by the theoretical physics divisions of the major scientific laboratories. Nevertheless it seems that fundamental Theoretical Physics and Mathematics will be amongst the last bastions to fall to the widespread advance of computers, and the most striking illustration is perhaps provided by the Institute for Advanced Study at Princeton itself, where the first computer seems temporarily to have been the last. Why should this be?

Perhaps we should look at how theoretical physicists actually live. They rely on infallible, cheap equipment such as pencils and paper, chalk, blackboard and libraries, and they like to work at any hour of the day or night. Periods of intense interaction with colleagues are followed by periods of solitary thought or writing. Theoreticians employ highly symbolic languages which are universally understood, and which at the same time can be freely extended to allow the introduction of new ideas. They use a wide variety of symbols and positional notation and constantly draw diagrams in order to explain their meaning.

Computers have so far been the antithesis of this way of life. They frequently break down, they are noisy, very expensive, and produce great quantities of unwanted paper. In the past they have mainly been localized in one place and have often provided a long turnround for even the simplest operations. Their character set has apparently been designed for writers of telegrams, and even high-level languages such as Fortran have been cumbersome compared to mathematics and difficult to change except by international committees while Algol 60 has never recognized the existence of complex numbers. Limited data storage has been available - we have already seen in §4 that it is typically 0.05% of what a physicist would normally use. In striking contrast to the shared literature of mathematics and physics are the innumerable programs and subroutines which are confined to individuals or to small groups of users.

Yet are these limitations inevitable, or are they simply a reflection of the fact that computers have been with us only a short while? We know that they can be connected to the universal automatic telephone network, so that by dialling or more easily just by pressing a button one can be connected to any computer in the country or ultimately, in the world. The commercial computer utility which I habitually use both from home and from the Culham Laboratory provides an instantaneous, interactive Fortran service 7 days a week, 23 hours a day and virtually never breaks down; certainly overwhelmingly less frequently than it is necessary to sharpen a pencil. Von Neumann was amongst the first to point out [61] that reliable performance can be achieved from basically unreliable components and there are now many computer networks that are almost fault-free. Hence in two respects at least the theoretician's ideal can now be achieved and at the laboratory we have coupled the same utility to a Tektronix 4010 display in order to produce cheap interactive graphics in a noise-free environment.

There seems no reason why all the other practical requirements of the theoretical physicist should not now equally readily be met, and it would be an interesting exercise to design an optimum combination of hardware, software and programming languages for a theoretical institute, a quarter of a century after the electronic digital computer was first introduced.

References

1. 'Methods in Computational Physics', ed. Alder, Fernbach & Killeen (and previously by Rotenberg), Vol.1 (1963) - Vol.10 (1971), Academic Press, New York.

2. 'Journal of Computational Physics', ed. Alder, Fernbach & Killeen (and previously by Rotenberg), Vol.1 (1966) - Vol.9 (1972), Academic Press, New York.

3. F H Harlow, 'Numerical Methods for Fluid Dynamics, an Annotated Bibliography', Los Alamos Report LA-4281 (1970).

4. 'Computers and their Role in the Physical Sciences', ed. Fernbach & Taub, Gordon & Breach, New York (1970).

5. See for example S J Aarseth, 'Perturbation Treatment of Close Binaries in the N-Body Problem', Astronomy & Astrophysics 9 64 (1970); M Hénon, 'Rates of Escape from Isolated Clusters with an Arbitrary Mass Distribution', loc.cit. 2 151 (1969), and papers on the 3-body problem and isolating integrals which may be traced from M Hénon loc.cit. 1 223 (1969), M Froeschlé loc.cit. 16 172 (1972).

6. For two recent examples see Z Barkat, J C Wheeler & J R Buchler, 'The Evolution of Growing Stellar Cores', Astrophys.Journ. 171 651 (1972); W K Rose & R L Smith 'Nova Hydrodynamics', Astrophys.Journ. 172 699 (1972).

7. 'Computer Physics Communications', ed. Burke, Vol.1 (1969) - Vol.3 (1970) North Holland, and the 'International Physics Program Library', Queen's University, Belfast.

8. K V Roberts, 'The Publication of Scientific Fortran Programs', Comp.Phys.Comm. 1 1 (1969).

9. J P Christiansen & R W Hockney, 'FOUR 67, a fast Fourier transform package', and 'DELSQPHI, a two-dimensional Poisson-solver program', Comp.Phys.Comm. 2 127, 139 (1971).

10. IPPS Conference on Computational Physics, Culham Laboratory, July 1969. Proceedings published as Culham Report CLM-CP(1969), 2 vols, HMSO London.

11. Second Annual Conference on Computational Physics, Imperial College, London, Sept.1970. Conference Digest published by Institute of Physics and the Physical Society, London.

12. 'The Impact of Computers on Physics', Inaugural Conference of the European Physical Society Interdivisional Computational Physics Group, CERN, Geneva, April 1972. Proceedings to be published in Computer Physics Communications.

13. 'CERN Courier', Vol.12 No.3 March 1972; special issue on Computers.

14. 'Computers as a Language of Physics', Seminar Course held at the International Centre for Theoretical Physics, Trieste, August 1971, Proceedings to be published by the IAEA.

15. H W Emmons, 'Critique of Numerical Modelling in Fluid Dynamics Phenomena', Annual Review of Fluid Mechanics, Vol.2, ed. M van Dyke, W G Vincenti & J V Wehausen (Palo Alto, California) pp.15-36.

16. P M Morse & H Feshbach, 'Methods of Theoretical Physics', Parts I & II, McGraw-Hill, New York (1953).

17. N J Zabusky, 'Solitons and Energy Transport in Non-linear Lattices', loc.cit. ref.12.

18. I Iben Jr, 'The Role of the Computer in Astrophysics', loc.cit. ref.4 pp.595-608 gives a general discussion on the use of computers in astrophysics.

19. See for example Chapter 6, 'Calculation of Stellar Structure', in 'Principles of Stellar Evolution and Nucleo Synthesis' by D G Clayton, McGraw Hill, New York (1968).

20. H C Thomas, 'Computer Programs in Astrophysics', loc.cit. ref.12 .

21. R Betchov & W O Criminale, 'Stability of Parallel Flows', Academic Press, New York (1967).

22. B Zacharov, 'Computer Systems', loc.cit. ref.12 .

23. J R Pasta, 'The Heuristic Value of Computers in Physics', loc.cit. ref.12.

24. G K Batchelor, 'Theory of Homogeneous Turbulence', Cambridge University Press (1953); G K Batchelor 'Proceedings of the International Symposium on High Speed Computing in Fluid Dynamics, Monterey (1968), published in Phys.Fluids Suppl.II 233 (1969). See also other papers in the same Proceedings.

25. R Kraichnan, Phys.Fluids 10 1417 (1967).

26. C E Leith, Phys.Fluids 11 671 (1968).

27. J G Charney, 'Meteorology', loc.cit. ref.12 .

28. H L Berk, C E Nielsen & K V Roberts, 'Phase Space Hydrodynamics of Equivalent Non-linear Systems: Experimental and Computational Observations', Phys.Fluids 13 980 (1970).

29. J P Christiansen & K V Roberts, Paper 40, loc.cit. ref.10. J P Christiansen, Culham Report CLM-R106(1970) & Culham Preprint CLM-P282 (1971). K V Roberts & J P Christiansen, 'A Survey of Computational Fluid Mechanics', loc.cit. ref.12.

30. B J Alder & W G Hoover, 'Numerical Statistical Mechanics', Chapter 4 of 'Physics of Simple Liquids', ed. Temperley, Rowlinson and Rushbrooke, North-Holland (1968). This contains references to previous work on the molecular dynamics method.

31. W W Wood, 'Monte Carlo Studies of Simple Liquid Models', Chapter 5, loc.cit. ref.30. This gives many references to earlier work.

32. E Fermi, J R Pasta & S M Ulam, 'Studies in Non-Linear Problems,I', Los Alamos Report LA-1940 (May 1955). Also in 'Collected Works of E Fermi', Vol.II, pp.978-988, University of Chicago Press (1965).

33. See the forthcoming review paper by R M Miura, 'The Korteweg-de Vries Equation: A Model Equation for Dispersive Waves', in 'Lectures on Non-Linear Wave Propagation', ed. Leibovich & Seebass (to be published).

34. R D Richtmyer & K W Morton, 'Difference Methods for Initial Value Problems', 2nd Ed. Interscience, New York (1967).

35. J von Neumann & R D Richtmyer, J.Appl.Phys. 21 232 (1950).

36. R F Christy, 'A Study of Pulsation in RR Lyrae Models', Astrophys.Journ. 144 108 (1966).

37. L Spitzer,Jr & M H Hart, 'Random Gravitational Encounters and the Evolution of Spherical Systems. I.Method & II.Models', Astrophys.Journ. 164 399 & 166 483 (1971). These papers describe a new method of calculation and compare this with previous methods; they provide an excellent illustration of the impact that computers can have on astrophysical thinking.

38. F Hohl, 'Numerical Experiments with a Disc of Stars', Astrophys.Journ. 168 343 (1971); R H Millar, K H Prendergast & W J Quirk, 'Numerical Experiments on Spiral Structures', Astrophys.Journ. 161 903 (1970).

39. C C Lin, C Yuan & F H Chu, Astrophys.Journ. 155 721 (1969).

40. G Kuo Petravic, M Petravic & K V Roberts, 'A 3-Dimensional Relativistic Computation for the Pulsar Magnetosphere', ESRIN Conference on Cosmic Plasma Physics, Frascati, September 1971, Proceedings edited by K Schindler, Plenum Press (1972).

41. J A Wheeler, 'Geometrodynamics', Academic Press, New York (1962).

42. 'Handbook of Mathematical Functions', ed. M Abramowitz & I A Stegun, National Bureau of Standards, republished Dover, New York (1965).

43. A Goldberg & J L Schwartz, 'Integration of the Schrödinger Equation in Imaginary Time', J.Comp.Phys. 1 433, 448 (1967).

44. R C Grimm & R G Storer, 'Monte Carlo Integration of Schrödinger's Equation', J.Comp.Phys. 7 134 (1971). This contains references to previous work.

45. S V Lowande, C A Jensen & H L Sahlin, 'Monte Carlo Integration of the Feynman Propagator in Imaginary Time', J.Comp.Phys. 3 416 (1969).

46. E Schrödinger, 'What is Life? The Physical Aspect of the Living Cell', Cambridge University Press (1945).

47. N Bohr, Z.Physik. 9 1 (1922).

48. D R Hartree, Proc.Camb.Phil.Soc. 21, 625 (1923).

49. R B Lindsay, J.Math. & Phys. 3, 191 (1924).

50. J C Slater, 'The Electronic Structure of Atoms - the Hartree-Fock Method and Correlation', Rev.Mod.Phys. 35, 484 (1963).

51. P G Burke, 'The Impact of Computers on Atomic Physics', loc.cit. ref.12 .

52. A R Davies, K Smith & K L Kwok, Comp.Phys.Comm. 3 (1972)

53. F J Smith, Invited papers at the VII Int.Conf. on Elec.Atom.Coll. (1971) (North-Holland).

54. A C Wahl, 'Chemistry from Computers; A New Instrument for the Experimentalist', Proceedings of the Fourth Conference on Numerical Simulation of Plasmas, held at Naval Research Laboratory, Washington D C, November 1970. Published by Office of Naval Research, Department of the Navy, July 1971.

55. 'Atoms to Molecules' (Film Series) by A C Wahl & V Blukic, McGraw Hill, New York (1969) & J.Chem.Educ. 45, 787 (1968).

56. S Cohen, 'The DELPHI-SPEAKEASY System. I. Overall Description', Comp.Phys.Comm. 2, 1 (1971)

57. E A Hylleraas, 'Reminiscences from Early Quantum Mechanics of Two-Electron Atoms', Rev.Mod.Phys. 35, 421 (1963).

58. Y Accad, C L Pekeris & B Schiff, Phys.Rev.A4 516 (1971); B Schiff, C L Pekeris & Y Accad, Phys.Rev.A4 885 (1971).

59. J A Campbell & A C Hearn, 'Symbolic Analysis of Feynman Diagrams by Computer', J.Comp.Phys. 5, 280 (1970). J R Chisholm, 'Algorithms for the γ-Algebra of Electromagnetic Form Factors', J.Comp.Phys. 8, 1 (1971).

60. K V Roberts & J P Boris, 'The Solution of Partial Differential Equations using a Symbolic Style of Algol', J.Comp.Phys. 8, 83 (1971); M Petravic, G Kuo Petravic & K V Roberts, 'Automatic Optimization of Symbolic Algol Programs, I. General Principles', J.Comp.Phys. (to be published).

61. J von Neumann, 'Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components', Automata Studies, Princeton University Press (1956). Reprinted in 'Collected Works', Vol.V, p.329, Pergamon Press, Oxford (1963).

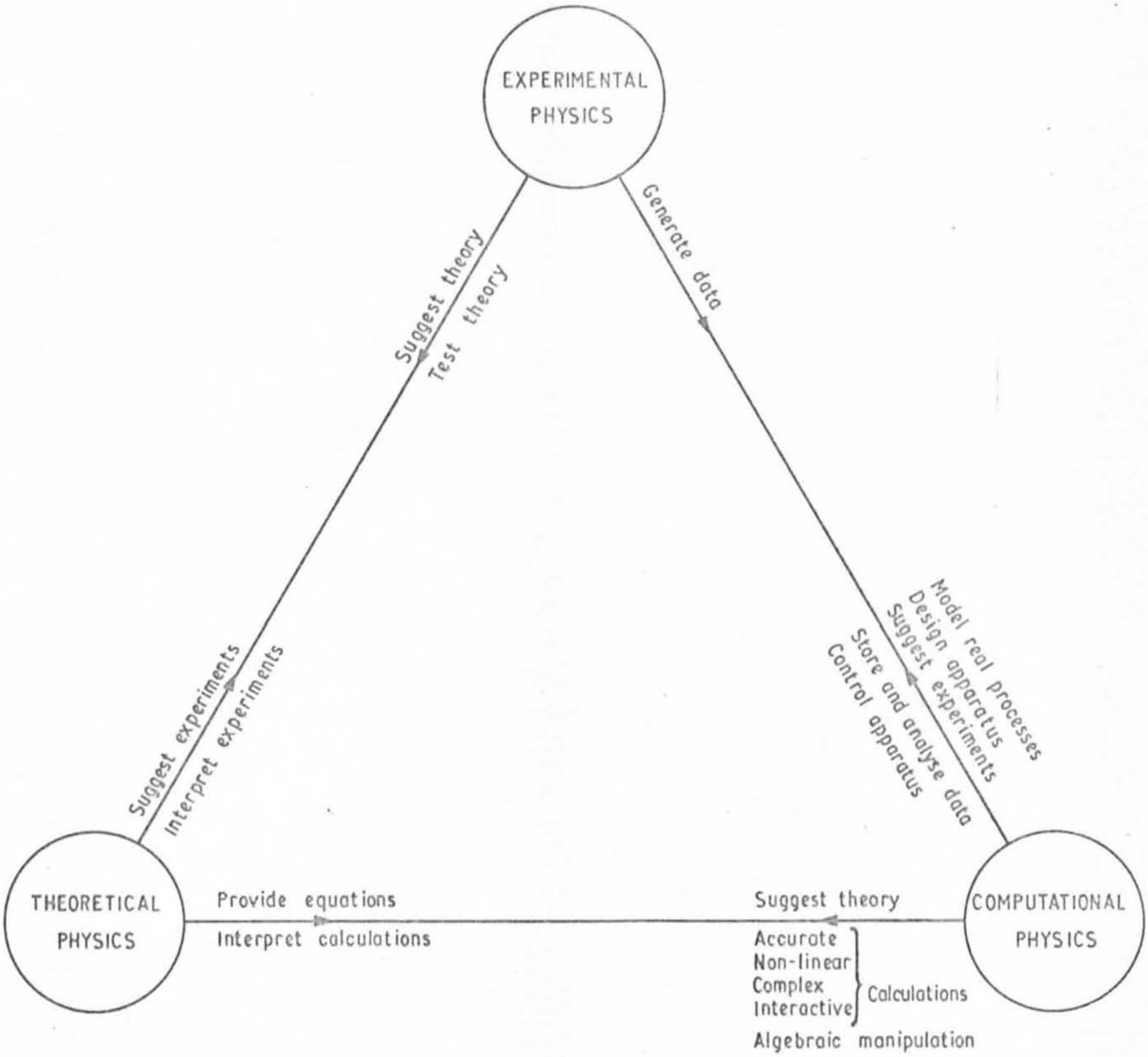


Figure I. Three approaches to physics problems.

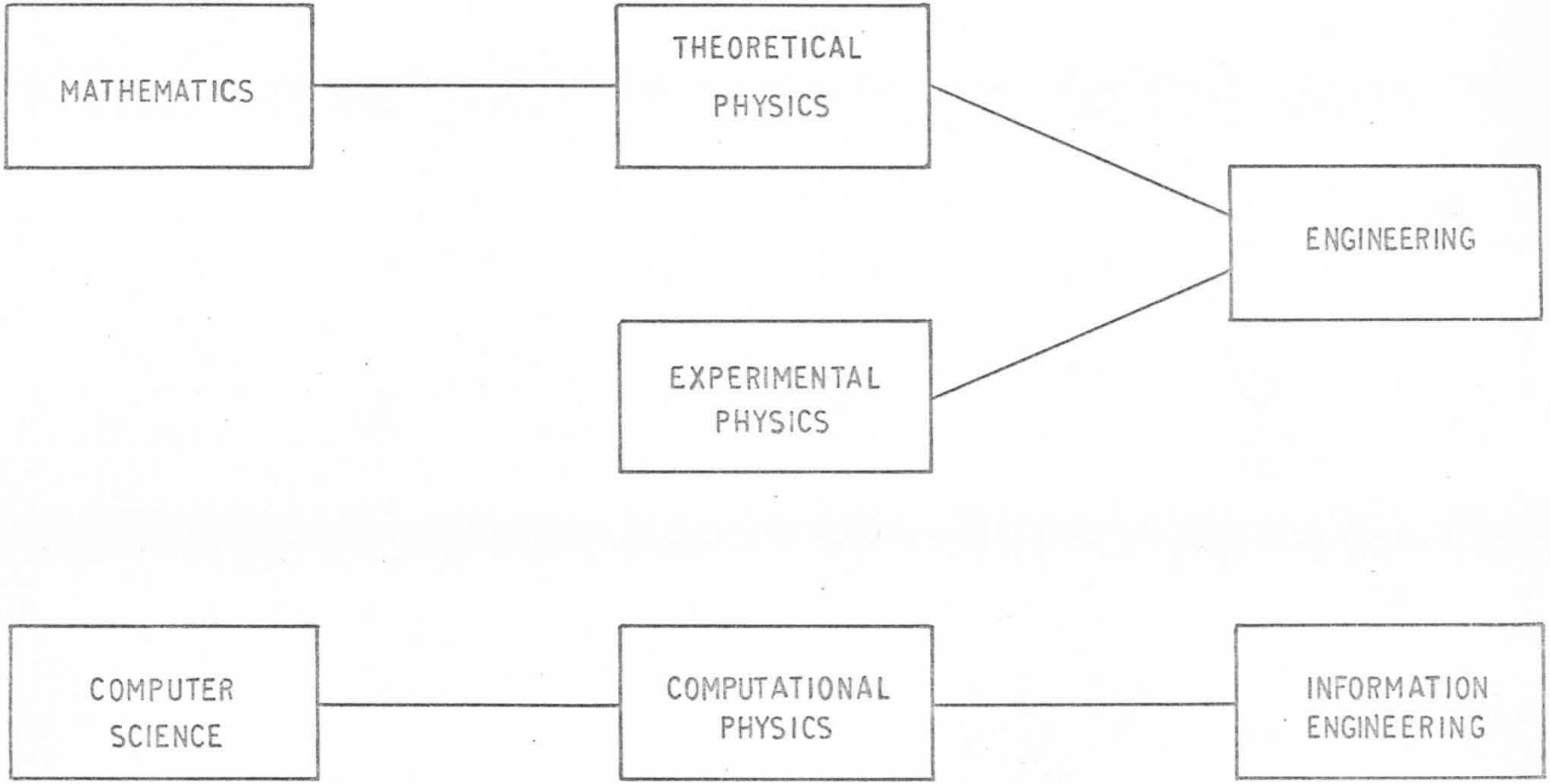


Figure 2. Relations between physics and other disciplines.

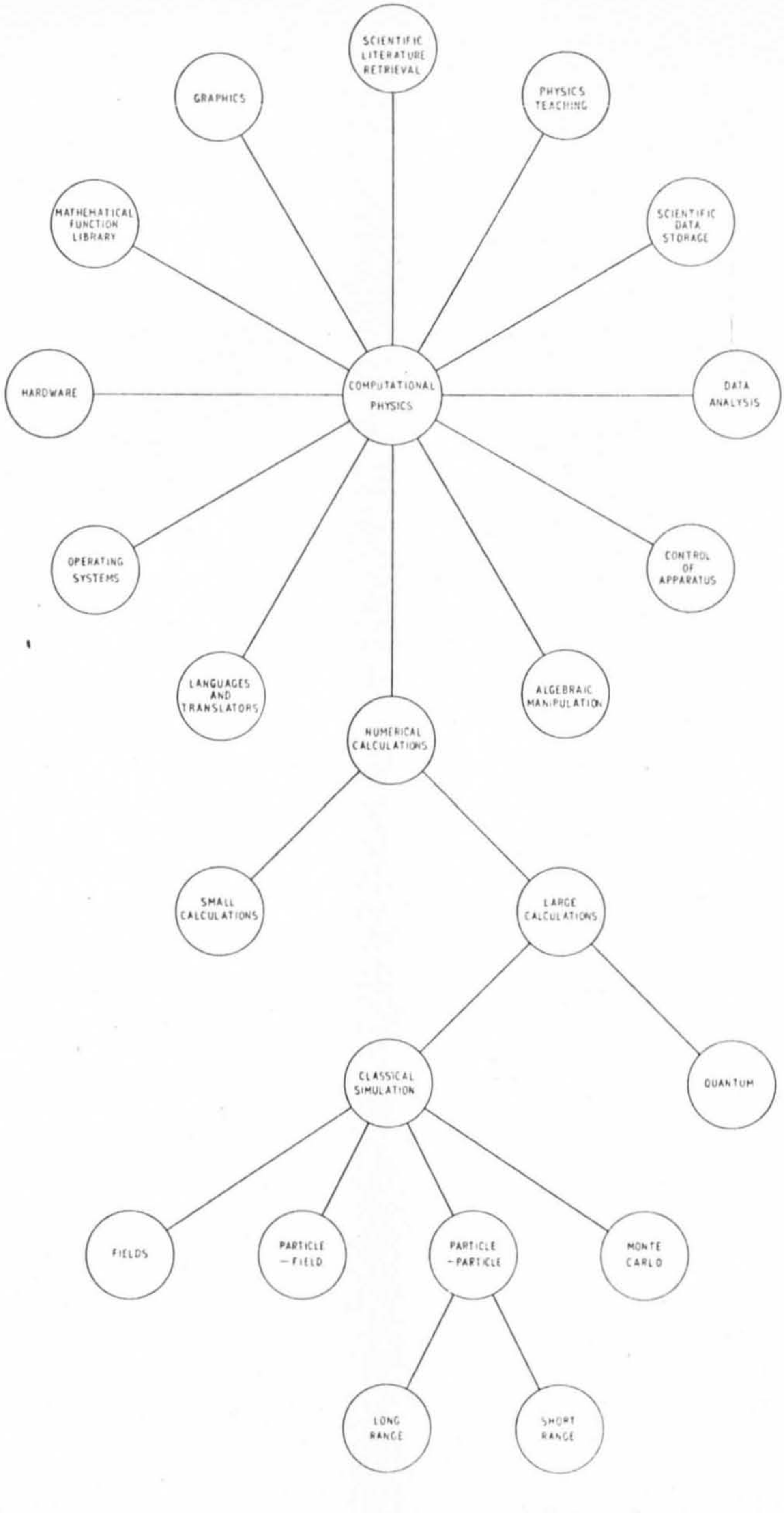


Fig.3 Some of the areas in which computing has an impact on physics.

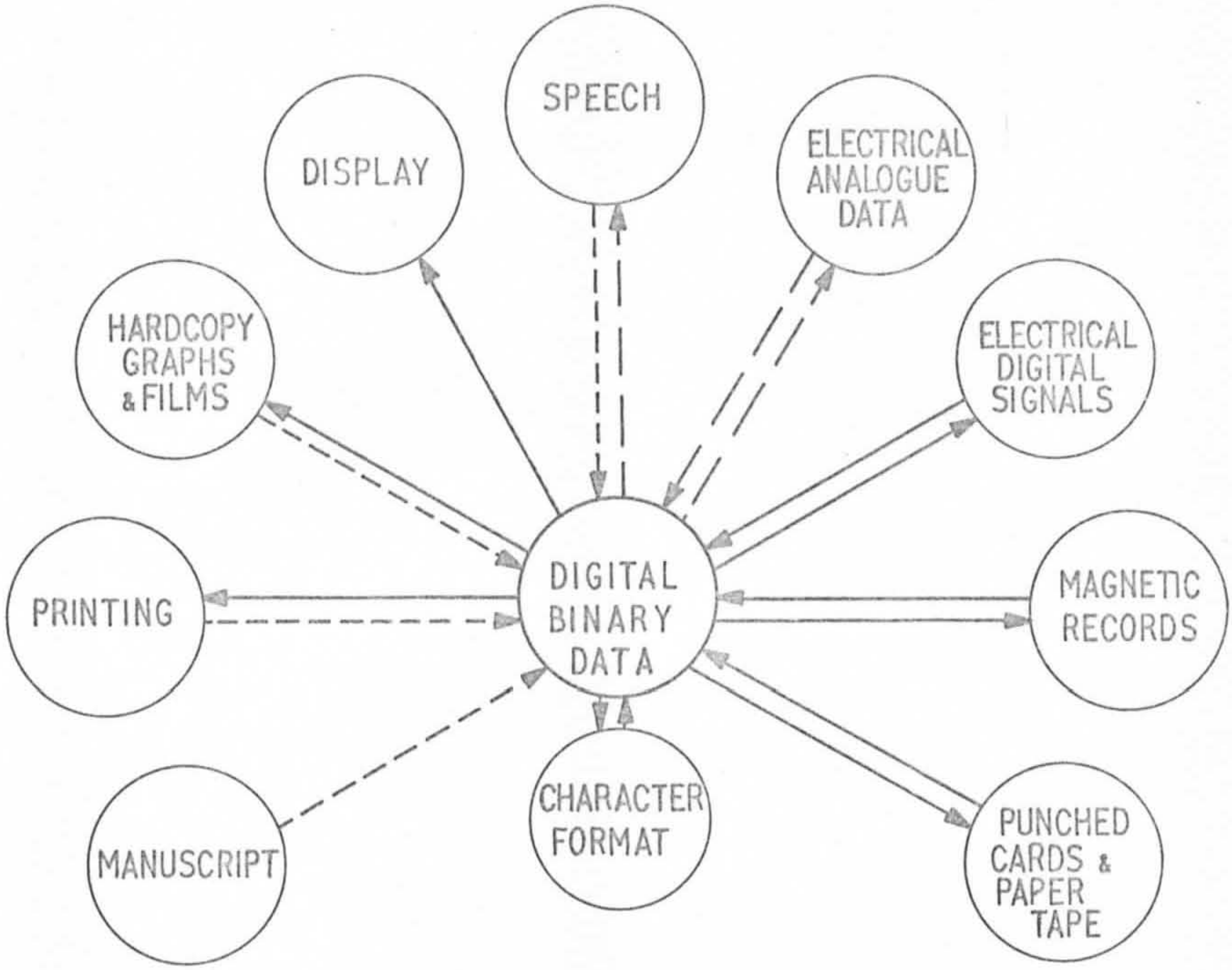
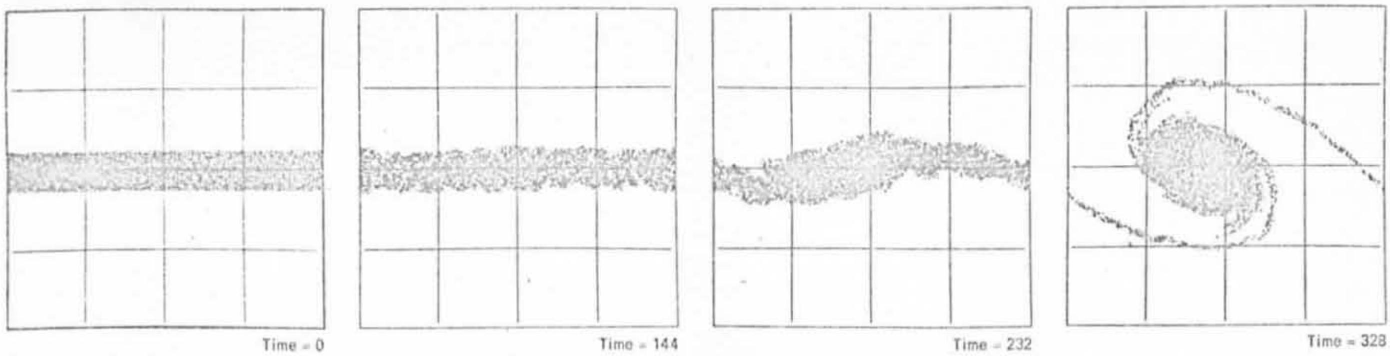
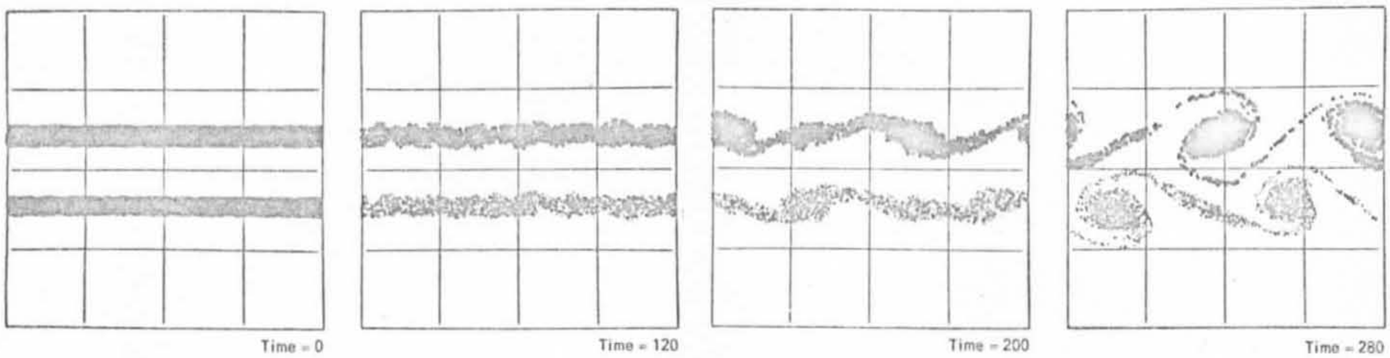


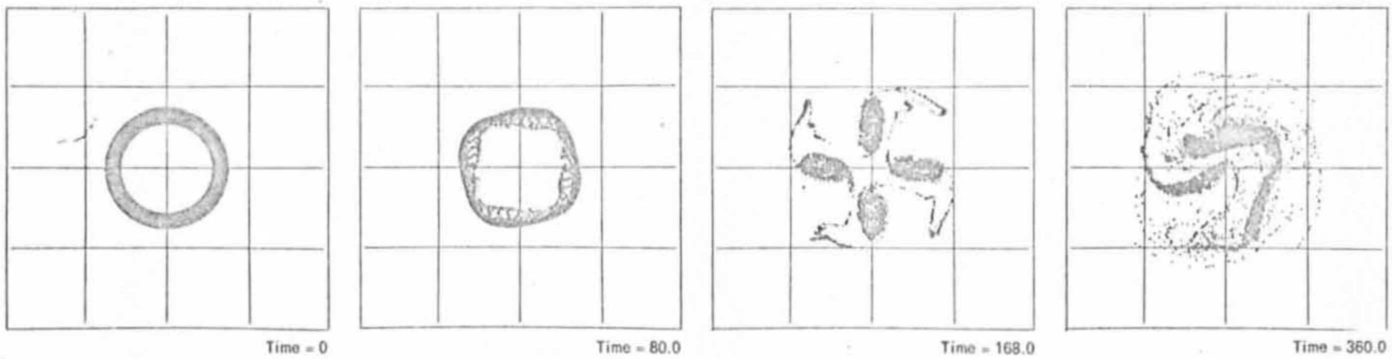
Fig.4 Data Transformations



(a) The onset of the Kelvin-Helmholtz instability



(b) The formation of the von Karman vortex street



(c) Sheared rotation. Diochotron type instability

Fig. 5. Some non-linear instabilities

LANGUAGE, LEARNING, AND MODELS OF THE MIND

George A. Miller

The Rockefeller University

and

The Institute for Advanced Study

What has been the impact of the computer on linguistics and psychology? Let me approach the question in an academic context.

In 1940, when I arrived at Harvard as a graduate student, I think that the Department of Psychology owned about three major pieces of equipment. I can remember two of them, and I add one to allow for faulty memory. I call them "major" because they must have cost about $1000. The laboratory consisted of several rooms plus an all-purpose "shop" for machine work, carpentry, electrical work, and painting. As for computers, there was a chap at Harvard playing around with relays and typewriters, but that was very remote from anything that went on in the Department of Psychology. At that time Harvard's Department of Psychology was one of the best in the country, perhaps in the world, so it is not unfair to use it as a benchmark for estimating subsequent developments.

We need a benchmark, for it is difficult even for those of us who have lived through it to remember how rapidly the face of science has changed in the last thirty years. World War II mobilized scientists of all disciplines, gave them urgent problems and the funds and facilities to pursue them. The success of that effort led to continued funding after the war, with consequent exponential growth in almost every dimension of the scientific enterprise -- in manpower, in budgets, in facilities, in technological applications, in publications, and even, I like to think, in good ideas. This growth was greatest in the best-established sciences, of course, but even young disciplines like psychology received as much support as they could intelligently use. Linguistics, being even younger as a science and still regarded by many people as one of the humanities, was less affected, but even there the repercussions of this national science policy were felt.

When one tries to understand the impact of the computer on such fields as psychology and linguistics, it is necessary somehow to factor out the recent changes attributable directly to computers from those attributable generally to increasing affluence. Such factorizations are difficult, if not impossible, for the computer is available only to those with sufficient resources to purchase or rent, staff, maintain, and program it. Today the psychology departments in every major university have one or more computers of their own, and often each laboratory in the department will have one or two. Linguistics departments do not actually operate their own machines, for linguistic research seems to require very large memories, but most universities have someone around who is competent in "computational linguistics" and who can be seen entering and leaving the computational laboratory at odd hours of day or night. This kind of facility, this kind of computational activity, costs manpower, time, and money; affluence is a necessary precondition. And providing the funds necessary to keep such installations humming has turned more than a few hairs gray on the heads of department chairmen, deans, and university presidents.

In order to appreciate what computers have done to psychology and linguistics it is necessary to consider particular cases, cases in which the effect of the computer is obvious, and to neglect the general changes in attitudes, styles of work, and ways of thinking that accompanied the computer revolution. It is my personal belief that the indirect effects of computers on the way we go about our science are the most important effects of all, but so much was happening on the scientific scene all at once that factorization is a questionable undertaking. I return to consider the general change in ambience later. First we need some specific examples of computer-induced research by way of background.

Psychology

Let me begin with psychology, since the effects there have been more profound than in linguistics and I know them better. There has been an historical trend for psychology to become more quantitative every year, a trend that could easily be documented by counting the variety and complexity of uses of statistical analysis of data appearing in psychological journals. This trend antedates the computer by half a century, so that, by the time computers became available, many psychologists already had problems worthy of them. Perhaps a typical example is the method of analyzing data that is called "factor analysis."

Factor analysis is a procedure for analyzing a matrix of correlations computed between the results of psychological tests taken by a group of individuals. A battery of tests is administered to a sample of people, their scores on each test are correlated with their scores on every other test, and the result is a correlation matrix. Factor analysis determines the rank r of this matrix, and represents the tests as points in an r-dimensional space; two tests that measure the same thing should fall near one another in this space. The coordinates of the space are then rotated until the coordinate values for each test seem to reflect a psychologically meaningful configuration -- meaningful in terms of the psychologist's intuitions about what traits or abilities the tests were testing.

Just to give some impression of what this method entails by way of computational labor, imagine that fifty different tests were administered to 200 individuals. I will neglect the time required to construct the tests and administer them, since the computer has not affected it. But scoring the tests would take about 1000 hours; to compute a Pearson product-moment correlation coefficient between two tests with 200 pairs of scores would take about two hours, so 1225 correlation coefficients means 2500 hours; factoring the matrix might take another 1000 hours; rotating the solution to obtain an optimal description might take another 500 hours. So a respectable factor analytic study represented an investment of about 5000 man-hours. If we assume a diligent researcher could devote 50 hours a week to this project, it would take him 100 weeks, or two years to finish it.

The amount of effort involved in such a study was well known, of course, and I personally held the belief that no psychology department would fail to grant a Ph.D. degree to any graduate student who successfully completed such a chore. But then the computer came on the scene.

Tests can now be scored by computer (indeed, they can even be administered by computer), and a variety of factor-analytic packages is available in every comp lab, so it is not impossible to finish off the whole business in about a week, once the production schedule is set up and running smoothely.[1]

The effect of the computer was not merely that the turn-around time for the researcher was reduced by a factor of about 100, thus making it possible to use the method experimentally, but statistical properties of the method could be studied and better computational techniques developed. A whole new variety of multi-dimensional techniques of analysis began to flower -- a development that we have certainly not seen the end of yet. The power and complexity of the hypotheses that could be tested by these and related methods of analysis of variance increased enormously. At least one veteran of this transitional period has referred to it as "the computer revolution in psychometrics."[2]

This is a familiar story, of course, repeated in every discipline that was ready to exploit the number-crunching power of the new machines. I suspect that in psychology the machines have put muscles on our analytic techniques that are more powerful than the ideas those techniques enable us to explore. But even if I am right, that is a cultural lag we will surely make up in a few more years.

The use of the new machines in information retrieval has had relatively less impact on psychology than has their use as calculators. However, the Psychological Abstracts have been printed by computer for several years. The American Psychological Association now makes available, thanks to support from the National Science Foundation, a remote console that enables subscribers to interrogate the Abstract's files.[3] This effort represents a modest attempt to cope with exponentially growing publications, and should be viewed in the general context of the library problem as a national emergency. In the practice of psychology, computers have been used to aid psychological counselors by providing automatic retrieval of educational, vocational, and occupational information.[4] No doubt other needs for information retrieval will be found in the years ahead.

An important application of computers in experimental psychology is their use in controlling the conditions of experimentation, both with animals and humans. This use not only permits more precise control of stimuli and measurement of responses, but has the advantage of permitting modifications in subsequent events contingent on the outcomes of earlier events. For example, if a test item evokes a response that indicates it was too easy, the computer can select a harder item next; if it was too hard, it can select an easier one. By such a strategy the computer can "track" the person's critical point from trial to trial in a manner that would be slow and clumsy for a human experimenter. This flexibility has practical implications for the use of computers as teaching machines as well as in the experimental laboratory.

A computer can also be used as part of a bio-feedback loop to display to a person some bodily process that he is normally unaware of -- his skin temperature, brain waves, heart beat, etc. -- and can administer rewards or penalties for specific changes in these processes. The burgeoning interest in bio-feedback in recent years does not necessarily involve computers, but they greatly increase the variety and flexibility of such research.

Calculation, retrieval, and control are relatively mundane ways to exploit computer power. Psychologists have contributed little to the state of the art that was not done previously, and perhaps better, in other branches of science and technology; it has been simply a matter of adapting tools of general importance to the particular needs of psychology. The application that has received the most discussion, however, and that many feel to be the most significant of all, is the use of computers to simulate the behavior of human beings.

Attempts to program machines to behave like intelligent organisms have had a tangled and complex history that I will not attempt to relate here.[5] But some major divisions of interest have emerged. For example, simulations based on neural networks and simulations based on cognitive processes are fundamentally different. Early simulations were largely attempts to move from random neural nets to more complex functional components. As it became clear that functional order did not emerge easily from total neural randomness, more recent work has tended to favor simulation based on cognitive processes whose realization at the neural level is indeterminate and largely irrelevant. Both kinds of simulation continue, but the more abstract cognitive approach seems to be of greater psychological interest.

There is an important and not always appreciated distinction between simulation and artificial intelligence. A scientist who simulates makes an explicit claim that his computer program performs its work in the same way a living organism does. Most scientists interested in artificial intelligence, on the other hand, merely want to get the job done as effectively as possible, regardless of how humans do it. They aspires to machines that are even better than we are. As it is sometimes expressed, if a worker in artificial intelligence wanted to build a new and better form of transportation, he would not begin with a walking machine just because it was a good simulation of living organisms, but would prefer to move directly to the more efficient wheel. Simulation should be of greater psychological interest, of course, but in recent years the two lines of attack have converged to a general search for principles sufficiently abstract to characterize any kind of information processing, living or non-living, so that there is really little difference between them now.[6]

And there are disagreements over the validation of a simulation. There is no problem here for those interested in artificial intelligence; either their machine does something intelligent or it doesn't. But if one claims to have simulated human behavior, then one must somehow test whether the machine and people behave similarly with respect to the functions one claims to simulate. Turing's "imitation game" is seldom an adequate test. Comparing computer traces with human introspections is more convincing, but hardly conclusive evidence that anything more than an analogy is involved. Productivity is a strenuous test; if simulation A and simulation B can be combined to produce a plausible simulation C for behavior that was not directly considered in creating either A or B, then both A and B inherit credibility from this extension. And if a simulation predicts behavioral phenomena that have never before been observed, then at least its usefulness is validated.

There seems to be a general impression that validating simulations is not really very important, since the work seems to progress quite vigorously in spite of the lack of any rationale for validation. That must mean that the advocates of simulation have, for better or worse, adopted the criterion of artificial intelligence -- either it works or it doesn't. But that criterion leaves in limbo the psychologically interesting claim that a simulation program is a theory of the behavior it simulates. Surely there must be some way to confirm or infirm our theories.

Increasing familiarity with computers and computer programming, plus the optimistic hope that programs are concrete actualizations of otherwise abstract theories, have led to an increasing frequency of flow-charts in the psychological literature. It is almost as if programming (usually in a rather rudimentary form) had provided a general language for describing psychological processes. And as the language of computing has come to be more widely used in stating theories than in programming computers, psychology has begun to lose its separate identity and become one of a large family of intellectual enterprises than are sometimes called the "computer and information sciences." The intelligent mind comes to be viewed as just another example of an information processing system, governed by the general laws that govern any such information processing system.

Attitudes toward this potential assimilation of (cognitive) psychology into the family of computer and information sciences differ. Some psychologists welcome it as a new and broader view of their universe. Some reject it, just as they would reject the claim that psychology must be a branch of physics because organisms obey the law of gravity. Those who have been thoroughly indoctrinated by a discipline-oriented education cannot get it out of their heads that psychology has a unique subject matter of its own, and that information processing is only part of the problem they face in gaining a deeper understanding of man and human nature. However, no objective reporter of the current scene could deny that psychological explanations are being phrased in the language of computer programming with ever increasing frequency. It forces one to be clearer as to what he is talking about. As R. W. Hamming has said, "Without a detailed description in some language that a machine can use there is no conviction that you know what you are talking about; with it there is the illusion you do."[7]

But now we have returned to the question of the computer's impact on our way of thinking about psychology. The real importance of the computer in psychology today is that it has created a new and pervasive state of mind. Indeed, I think it is appropriate to say that for most psychologists "the computer" <u>is</u> a state of mind, not a piece of machinery. There is more to be said about that state of mind, but first we should consider briefly the rather different role that computers have played in the field of linguistics.

Linguistics

I should confess that my membership in the Linguistic Society of America dates only from 1950, and my credentials as a linguist earn me only the status of diletante 3rd class. As an interested outsider, however, it is my opinion that the computer, as a machine, has had essentially no impact at all on the field of linguistics. The computer as a state of mind, however, has had some effects. Let me try to support these judgments.

From the earliest days of computer technology it has been obvious that these machines were not merely calculators designed to deal with numerical symbols, but that they were general symbol-processing devices, as receptive to the symbols of language as to the symbols of mathematics. A word is just a "number" spelled in base-26 arithmetic, and a sentence is just a string of those "numbers." The use of computers to process natural language was an obvious consequence of this broader conception of computation, and the visionary fathers of this field were quick to recognize the possibilities. In order to dramatize this new conception of the computer, in order to arouse interest and support for this broader view, they sometimes gave examples of what computers could do "in principle" that ran well beyond what anyone really knew how to do. Their propaganda succeeded, however, and a number of large projects in the processing of natural language by computer were initiated.

Some of the possibilities that were foreseen were: (1) mechanical translation; (2) automatic information retrieval; perhaps as part of a (3) question-answering system; (4) computer recognition of speech; (5) computer generation of speech. The usefulness of such devices, if we had them, is reasonably obvious, so the suggestion that the computer had put them within our technological reach stirred considerable interest and activity.

Unfortunately, however, these early promises proved to be blank checks written against a nearly empty account. The difficulties of automatically processing language were far greater than the original enthusiasts had imagined. Mechanical translation, the first and probably the most generously supported of these various projects, provides the clearest example. MT was a very poor place to begin, since an explicit, machine-interpretable analysis of two languages (at least) was required at a time when we didn't know how to make such an analysis of even one language. Several projects underwent a parallel history: there was an initial stage of computerizing the lexicons, with much fuss about efficient algorithms for retrieval, onto which was tacked a rather minimal syntactic component for smoothing out the results of word-for-word substitutions. The first tests were generally encouraging, giving translations that could be deciphered about 80% correctly by a patient and interested reader. More funds were poured in to support the effort to improve the syntax and clean up the obscure 20% that failed. But then deeper difficulties began to appear, and it became increasingly obvious that syntax was going to be very difficult to explain to a computer. Various pragmatic attempts to patch and mend the existing systems were judged unsatisfactory. In 1966 a blue-ribbon committee of the National Academy of Sciences made official the growing impression that mechanical translation was premature, given our level of syntactic sophistication and the availability of human translators to do the important work more accurately and less expensively.[6] The work still continues, but on a much reduced scale.

Linguists who had been involved in this affair and who retained their interest in mechanical linguistics concluded that they had to go back to the drawing board and develop a better way to deal with syntactic analysis, but this time in the context of a single language. With respect to computers in linguistics, the result was the development and refinement of a series of "parsing programs" based on various generative theories of syntax that were emerging at this time, the most influential of which was that described by Noam Chomsky.[9] I think it is fair to say that at the present time we do know rather well how to parse sentences by machine. There are still several competing systems, but I think the most effective has been the "augmented transition network" approach described by W. A. Woods, which is the grandson of a program developed in Edinburgh by J. P. Thorne, who thought he was implementing Chomsky's theories in mechanical form.[10] As described by Woods, however, augmented transition networks can implement any kind of grammar a theorist believes in, so it has no direct implications about the validity of Chomsky's approach to grammar relative to any other at the present time.

Parsing programs have probably been most extensively exploited in the development of question-answering programs which enable a user to address natural language requests to a large data base -- medical records, airline schedules, baseball scores, or what have you. According to R. F. Simmons, the first generation of these systems ran into difficulties reminiscent of the mechanical translation debacle, even though most of the syntactic problems seemed to have been solved (or avoided by limitations on the syntactic or logical forms of questions that could be asked). This time, however, the diagnosis was that the residual difficulties resulted from a lack of any adequate semantic theory.[11] At the present time, therefore, the considerable ingenuity of these computer-linguists is being devoted to semantic analysis, with results that are interesting both to linguists and psychologists. At the moment, however, the exploration of computerized semantics is still in that initial stage of enthusiasm that usually means we have not yet bumped our heads against the real problem. The next phase will see an effort to formulate for computers the conceptual knowledge available to a human language user -- which will bring the whole enterprise squarely into one of the most difficult and least understood areas of human psychology.

Automatic language processing has contributed to linguistic theory only by posing specific questions that linguists have tried to answer, and by providing a source of funds to support linguists while they considered such questions. That influence is not inconsiderable, of course, but it is rather indirect. Who is to say that linguists would not have pursued essentially the same course even if automatic language processing had never been attempted? The problems have always been there, and modern formalisms have made the new attack possible. Once again, factorization is difficult.

One exception to this general conclusion that the computer has contributed relatively little to linguistic science has occurred in the study of phonetics. There is, of course, a long history of attempts to use machines to analyze and synthesize the sounds of speech, so when the computer arrived the stage was set to incorporate acoustic and phonetic theories directly into computer programs. For example, the computer can provide a dynamic simulation of the vocal tract, varying the parameters of the tract according to measurements made from X-ray pictures of speakers and correlating the output with acoustic analyses of human speech. One of the most imaginative applications of this technology has been to the study of the vocal tracts of other primates, leading to the conclusion that they could not produce the variety of vocalizations necessary for human speech.[12]

The most important result of this analytic work with computers has been to provide a physical and physiological foundation for what phonologists have called "distinctive features." These studies of the relations between articulatory configurations and the resulting acoustic signals indicate that there are preferred ranges of articulatory gestures, ranges which give rise to acoustic outputs with attributes that are distinctively different and relatively insensitive to minor perturbations in articulation.[13] Moreover, the perceptual mechanism responds selectively to these unique properties, so there is in some sense an optimal match between productive and receptive systems. During natural speech, of course, these ideal acoustic characteristics are modified by context, so that a listener must rely on secondary cues and on his general knowledge of the language in order to interpret the speech signal he hears. But the basic ground-plan of the speech system seems to be intelligible in scientific terms for the first time.

In phonetics, therefore, the computer has been an important tool in advancing research of a fundamental character. This exception proves the rule; it makes obvious what it means to say that no comparable impact of computer technology on linguistic theory can yet be discerned in the fields of syntax or semantics. It seems to be the case that the computer is most productive in areas where a considerable foundation of theory based on previous research already exists. If this basic theory is lacking and is invented solely for the purposes of computer application, the results are usually disappointing. This conclusion could probably be supported by workers in other fields.

Someone who wished to disagree with my judgment about the contribution of computers to linguistics could well take the following line. One of the most revolutionary changes in the way linguistic theories are stated has resulted from Chomsky's introduction of explicit formulas for writing syntactic rules.[9] A Chomsky grammar, for example, looks very much like a computer program. Indeed, the rewriting rules of the immediate constituent component of the grammar are borrowed explicitly from E. L. Post, who introduced them in order to discuss metamathematical problems of computability. Chomsky's formulations of syntax seem to lend themselves directly to computer implementation. Surely this whole development should count as a major impact of computers on linguistics.

The rebuttal is simple enough. Post's work antedates the computer revolution by many years. Chomsky's efforts to provide an integration of formal logic and Zelig Harris's ideas of transformational grammar were not stimulated by any interest in computers or mechanical translation. The theory of programming languages has borrowed far more from Chomsky than he has from it. There is every reason to believe that linguistic science would have progressed in much the same way it has if computing machines had never existed.

In a larger sense, however, such a critic would be correct -- in spirit if not in detail. Without the computer revolution and the state of mind that the computer metaphor induces, there would have been far less interest in Chomsky's work, fewer brilliant students to defend and criticize it, and less support for research and university appointments in this brand of linguistic theory. Both the computer and developments in formal linguistic theory are manifestations of our general intellectual climate. Although neither "causes" much to happen to the other, both contribute to a _Zeitgeist_ preoccupied with the mechanics of symbol manipulation.

The Computer Metaphor

This state of mind is more pervasive and more influential than the computer, which is, after all, merely the most tangible expression of this preoccupation. As such, the computer has become something more than a machine; it has become the symbol for one of the great triumphs of the 20-th century mind. In order to keep the distinction straight between the computer-as-machine and the computer-as-symbol, I have fallen into the habit of referring to them as the "computer" and the "computer metaphor," respectively. The computer metaphor has had a far greater impact on both psychology and linguistics than has the computer itself.

What is the "computer metaphor?" The most sweeping form of it that I have heard was provided by a highly skilled systems programmer who once explained to me that the universe is a computer and God is its programmer; God is still trying to debug the program, and it is our duty to help Him. Each of the various fields of science is attempting to understand one of the subroutines; each of the various humanities is studying some trace of the computer's operation. The metaphor was developed in considerable detail, until it became apparent that he was propounding a complete, intricate, and compelling metaphysical system.

The claim that the universe is a computer is what some students of metaphor like to call a "root metaphor." The usual kind of metaphor that literary critics analyze is a "descriptive metaphor;" a particular word or phrase is given an extended interpretation in order to describe some novel conceptualization for which the language has no adequate symbol.[14] Both kinds of metaphor have played important roles in the history of science; it is sometimes difficult to tell just where and when a metaphor stops being metaphorical and starts being accepted as a theory. The root metaphors that "the universe is a machine" and that "the universe is mathematical" are certainly very old and honorable ideas in Western science, so I probably should not have been surprised when they were combined by my programming friend. Within the frame of reference provided by such a metaphor, a great variety of specific theories are possible for specific phenomena.

It all begins, of course, with the theory of computability, and the conclusion of Turing and Church and others that any operations that can be described explicitly can be performed by a universal Turing machine. From which it seems to follow that the only reason something cannot be done by a universal Turing machine is that we don't understand it. When we do understand it, then it too will fall within the machine's domain. Given this interpretation of what "understanding" consists of, any attempt to suggest counterexamples becomes merely a confession of ignorance or, if one persists in claiming that he can understand something he cannot describe explicitly, one becomes a prototypical member of that class of people known as mystics.

In defense of adopting the computer metaphor in psychology, Newell and Simon have recently written that:

> "An abstract concept of an information processing system has emerged with the development of the digital computer.... The various features that make the digital computer seem machinelike -- its fast arithmetic, its simply ordered memory, its construction by means of binary elements -- all have faded in the search for the essential....With a model of an information processing system, it becomes meaningful to try to represent in some detail a particular man at work on a particular task. Such a representation is no metaphor, but

a precise symbolic model on the basis of which pertinent specific aspects of the man's problem solving behavior can be calculated."[15]

This passage makes explicit what many psychologists have come to take for granted in recent years, namely, that men and computers are merely two different species of a more abstract genus called "information processing systems." The concepts that describe abstract information processing systems must, perforce, describe any particular examples of such systems.

There is a subtle line, however, between saying that, insofar as men act as information processing systems, their behavior can be described in terms known to apply to all such systems, animate or inanimate, and saying that men are <u>nothing but</u> information processing systems. The "nothing but" of reductionism seems to creep in in direct proportion to the extent to which a scientist becomes confused about the difference between his metaphor and the part of the world he uses his metaphor to describe.

At first glance, such confusion would seem relatively harmless, since if there are any practical differences between the metaphor and the reality, they will eventually emerge when a theory based on that metaphor fails to fit the observed facts. Historians of science, however, are not agreed that either metaphors or theories are so simply evaluated. A scientist firmly in the grip of some root metaphor will see any discrepancies between a theory based on that metaphor and the observed "facts" as a challenge to revise the details of his theory, not as evidence that a completely new metaphor is required.

To one caught by the metaphor, therefore, it is very difficult to understand why other people, presumably intelligent enough in other matters, should persist in denouncing the computer metaphor as an obscene vulgarization of all that is best in human nature. They see man as something more than an information processing system; they fear the computer metaphor may cause us to overlook that something more. If the metaphor is too successful, it might make over man's image of himself. Elting Morison has written:

> "...a machine, any machine, if left to itself, tend to establish its own conditions, to create its own environment and draw men into it. Since a machine, any machine, is designed to do only a part of what a whole man can do, it tends to wear down those parts of a man that are not included in the design."[16]

Will the universal Turing machine "wear down those parts of a man that are not included in the design?" From within the metaphor comes the response: One who fears so should try to say what those parts are, for if he can describe them, they can be included, and if he cannot, they are irrelevant. And from outside comes back the insistent question: Irrelevant to society or irrelevant to science?

After a few rounds back and forth at this level of abstraction a practical man loses interest in what may or may not be possible in principle. The fact is that for centuries psychological speculations were either incurably vague or obviously simplistic. Computers and computer languages provide a way to talk about psychology that is not vague, and while it still may be too simple, is at least orders of magnitude more complex than was feasible before. Even linguists have recognized the value of more explicit, detailed complexity in their theories. Ultimate truth may not yet be within our reach, but talking about men as information processing systems is certainly an improvement over the way scientists used to talk about them.

A practical man, however, once he gives up arguments about what is possible or impossible in principle, is likely to want to take a hard look at what is actually being done with the information processing model of man. Leaving aside any claims as to how much better things are now than they were before we had this new metaphor, he would insist on asking: how good are things now? If I may place myself momentarily in the shoes of the practical man, I think he would find some reason to doubt whether we are really launched at last on the high road to psychological or linguistic wisdom.

Rather than try to document these doubts by a critique of what existing systems can actually accomplish by way of pattern recognition, inductive reasoning, game playing, decision making, information retrieval, question answering, sentence parsing, mechanical translation, computerized teaching -- the value of any comment on an existing system has a half-life as brief as the current systems themselves -- I would like to consider a problem that I have thought about in some detail, namely, the problem of meaning. Meaning is a phenomenon of considerable importance to both psychologists and linguists, and one with a long history of philosophical discussion as well. By asking how the information processing metaphor has dealt with this problem, we can perhaps gain some rough estimate of how far the metaphor has brought us in practice, and how far remains to go in principle.

Meaning

Let me take up the story around 1950. At that time American psychology was very much under the spell of behavioristic dogma, which either rejected meaning as a serious topic for scientific investigation, or attempted to capture it en passant by some characterization in terms of associations acquired during an organism's past history of behavior and reinforcement. And at that time American linguistics was dominated by structuralism, which either rejected meaning as a serious topic for scientific investigation, or attempted to capture it en passant by some characterization in terms of the substitutability of one symbol for another in grammatical structures. What a layman would ordinarily understand by "meaning" was to mentalistic, too subjective, too invisible, too ephemeral for a positivistic science. The fact that this attitude left a gaping hole at the very center of what the layman would have thought these sciences were about was considered the result of a regretable lack of insight on his part into the way science progresses.

Around 1950, however, psychological theorists began to hear about computers and computer programming. One of the major results of this technological innovation was to provide a vastly enlarged conception of what a machine could be and do. Thus, with no sacrifice of scientific objectivity, psychologists were able to construct much more complicated theories and to cast them in the form of computer programs. Given the metaphor of the organism as an information processing system, therefore, various mentalistic concepts were revived and reconsidered in the light of this theoretical liberation. Intention was interpreted in terms of feedback and control, attention was seen as related to serial processing, memory was discussed in terms of storage and retrieval of information, thought was computation, and so on. So it was probably inevitable that meaning would also be reinterpreted in the light of the computer revolution.

Among psychologists we can distinguish two lines of attack on meaning during this period. One line was simply a continuation of the traditional associative theory of meaning, but now crystallized in terms of associations between items of information stored in the computer's memory, with "pointers" connecting one item to all its associates. Learning, in this view, is simply the procedure of depositing information in the memory and gradually connecting it in appropriate ways, based on experience and reinforcement, with other information similarly stored.

The second line was more radical, however, and deserves a special name to distinguish it from the simpler associative approach. Historically, it should be classified as an operational theory of meaning, but since that term has so many philosophical connotations, I will call it the "procedural" theory of meaning instead. According to this view, a meaning is a procedure that is associated with a stimulus and serves to make that stimulus a symbol. The meaning of the symbol "dog," for example, is the operational procedure for determining whether a given object is a dog -- by looking at it, counting its legs, hearing it bark, noticing its teeth, tail, fur, shape of body, and so on.[17]

It is relatively easy to appreciate how this procedural theory of meaning might emerge from experience with computers. When you tell a computer to do something -- to "add four and six," for example -- what is the meaning of the instruction "add?" The meaning is what you want the computer to do, that is to say, it is the program of operations that the computer is expected to execute. Analogously, when we say to a person, "Give me the book," the meaning of "give" is the program of instructions that the person is expected to execute in reponse. If the person does not have such a routine for "give" stored in his memory, then he does not know the meaning of that word.

Computer scientists recognized very early that the distinction between data and programs-that-operate-on-the-data is blurred and can be largely ignored; any particular bit of information can serve as data in one computational context or as instruction in another. A computer's knowledge of the world is not necessarily divided into facts and things-to-do-with-facts; both facts and things-to-do could be stored in the same format as routines that could be either examined or executed on demand. And if a computer's memories could be so dynamically organized, why not people's?

The procedural theory of meaning was an important advance over the traditional psychological definition. Superficially, it resembles a behavioristic approach, since it is framed in terms of what the system is expected to do when a stimulus occurs. But at a deeper level there are important differences. Note, for example, that the meaning of "dog" is not the particular dog that the sound may designate at the moment, nor is it our perception of that dog, or the class of objects that "dog" can refer to, or the speaker's intention in uttering "dog," or the set of environmental conditions or previous experiences that caused the speaker to use the utterance, or a mental image of some dog or other, or all the other words that "dog" makes us think of, or even a dictionary definition of "dog." The meaning of "dog" is the set of operational procedures that we must perform in order to verify that some object is a dog. The procedural definition was like a breath of fresh air in a very stale room.

So things were very much better after the information processing metaphor was adopted than they were before. But that is not the question our practical man asked us. How good is the procedural theory of meaning?

In recounting this history I have neglected the philosophers, who have had a much more sophisticated variety of views about meaning than have either psychologists or linguists. But I cannot conceal the fact that operational theories of meaning have received considerable philosophical attention. Charles Saunders Pierce founded his pragmatic philosophy on an operational conception of meaning, and Percy Bridgman, the physicist and philosopher of science, made operationism familiar even to many psychologists during the 1930s and 40s. As the idea emerged in the writings of the logical positivists, to know the meaning of any statement was to know the conditions under which it would be true. If there are no conditions under which you could test the truth or falsehood of a statement, then it was considered to be meaningless. This view -- that meaning depends on verifiability -- is obviously closely related to the procedural theory of meaning I have just described as growing out of the information processing approach.

Now, the philosophers discovered that one of the difficulties that theories of this sort encounter is that it seems to be necessary to understand the meaning of a symbol <u>before</u> you know how to verify it or know what routines it calls for. Knowing what to do is certainly part of the meaning, but it cannot be the whole of it. And so even the logical positivists began to distinguish between the intensional meaning that you had to understand first and the extensional meaning that had to do with how you could verify it. This situation is still unsatisfactory, however, since most of the things we hear we never bother to verify.

In addition to routines for recognizing what things are labeled "dog" and what operations are to be performed in response to words like "add" or "give," therefore, we need something more that might be characterized as the concepts these words express.

Perhaps the best way to indicate the need for something more is to recall how computers deal with language. In recent years computers have been programmed to answer questions about particular bodies of information, and to do so in language closely approximating the ordinary language of everyday discourse. In order to do this, the computer must be able to parse the question, to extract from it a program of operations for retrieving information from memory and transforming it logically into a form appropriate as an answer, then to phrase that transformed information in a grammatical sentence that the questioner can understand. In mentalistic terms, the computer "understands" the question, "remembers" relevant information, "thinks" about the information in order to formulate an answer, and then "expresses" the answer in grammatical sentences. Proponents of the procedural theory of meaning would claim that the computer is doing just what a person would do in answering such questions, and that there is nothing more to meaning than the operational procedures that the computer is performing.

In evaluating such claims, however, it is necessary to remember that the question-answering systems developed to date deal with sharply restricted domains of information involving relatively limited vocabularies -- such things as the scores of sporting events, or airline schedules, or military logistics, or business inventories, or the positions and movements of a small set of blocks, or the titles listed in a library, and so on. Each program deals with one variety of question. However, since a person can deal with all these different realms of information, before we can be sure that the computers are doing exactly the same thing we are, we should consider what will happen when a computer's program and memory are expanded to include the much greater variety of information, terminology, and information processing that a person can combine.

Of course, there is nothing in principle that prevents us from simply adding more and more memory to the computer and giving it more and more procedures and more and more information. But simply combining two existing programs without considering their potential interactions would be inefficient and uninstructive. Twenty mutually independent programs wrapped up in one would be a gigantic monument to folly, and even a hundred would not converge toward the kind of information processing system exemplified by an intelligent human being. New principles and new systems will be required in order for computers to cope with the magnitude and variety of information that people can cope with. Something more -- something very like our own conceptual organization of that information -- will be required in addition to the operational procedures. And then, if I am right, we will recognize that the meaning of any symbol is given by its position in this conceptual organization, and that the procedural operations are merely ways of getting access to that position. Notice that I am not claiming that we can do or know anything that computers cannot do or know, but rather that at the present time we are still unable to characterize what we do and know in such a way that we can incorporate all of it into a computer program.

In order to indicate how the relation between conceptual organization and procedural operations might be conceived, it will help to shift the realm of discussion from computer science to taxonomic biology. Students of classification in general, and taxonomists in particular, draw a sharp distinction between what they call a "key," which is used to identify particular specimens, and the theory or conceptual basis for the key.[18] The theory of evolution, for example, might provide a basis for categorizing certain kinds of animals in a particular way, based on their geographical distribution, cross-fertility, innate behavior, fossil relics of different ages, and various other kinds of information. Once this information is conceptualized in some coherent theory and the appropriate categories are agreed on, it is a relatively simple matter to construct a procedure for identifying membership in those categories on the basis of whatever features are most convenient and reliable for distinguishing among them, and to arrange those procedures in a hierarchical sequence of questions so that a biologist can, by answering each question in turn, be led rather quickly to the correct label to apply to any particular specimen. This identification procedure is the key. It is based on a theory, but it is not the same kind of thing as a theory. The identification key is simply a convenient operational procedure for getting access to the theory, for locating the specimen in its correct conceptual position with respect to all the other categories of the theory.

What I would like to suggest, therefore, is that the operational procedures required to apply the label "dog" to a particular object, or to respond correctly to a request involving the verb "give," bear much the same relation to conceptual meanings as a taxonomist's key bears to the biological concepts underlying the key. It might be possible, for example, to program a computer to perform all the tests involved in a taxonomic key, so that the computer would behave as if it understood a great deal about biology, but unless the computer were also given the conceptual organization underlying that key, it would not "understand" what it was doing in anything resembling the same sense that a human taxonomist would understand it. A computer might be able to identify "man" by testing for featherless bipeds, but the conceptual meaning of "man" is certainly not captured by this identification routine.

In order to make these ideas more concrete, we need some examples, and the simplest example that is well understood is in the realm of basic color names. It is a familiar observation that children acquire the names for the colors rather late, considering their perceptual salience and the children's mastery of other nouns and adjectives. The typical developmental picture is that a child will learn to use "black" and "white" correctly rather early, but the names for the chromatic colors will come many months later and then will all appear in the child's speech over a relatively short period of time. My contention is that the child must develop a concept of color in terms of which each particular color label can take its place relative to all the others. Before that concept is attained, the boundaries between the colors cannot be understood; after the color concept is attained, the labels fit naturally together into a lexical field that covers the conceptual field like a mosaic.

The concept of color shared by all speakers of English is very close to Ewald Hering's theory of opponent colors,[19] that is to say, the conceptual field is organized in terms of a red-green opposition and a yellow-blue opposition, and these provide coordinates within which other colors can be arranged. This concept is organized, therefore, around four landmark colors, red, green, yellow, and blue, with gray at the center. Note particularly that this is a concept -- most people never really <u>see</u> the colors all laid out in a spatial pattern before they learn to name them. Moreover, people who speak a different language having fewer color terms may have a very different conception of color.

Now, given this concept, it is a relatively simple matter to devise a key based on judgments of the relative amounts of the landmark colors that we perceive in any particular sample. The key might ask such questions as: "Is there more red than blue?" and "Is there more red than yellow?" and "Is there more red than gray?" and so on. Depending on the answers, the key would specify a label that could be applied to the particular sample. If red and yellow were about equal, but there was more red and yellow than anything else, then the key would tell us to assign the label "orange" to that sample. And so on. The key would thus specify certain operational procedures that would be involved in responding correctly to the request to "Give me the red one," but those operational procedures would not be the <u>meaning</u> of "red." The meaning of "red" can only be understood in terms of its boundary relations to all the other color labels that span this particular dimension of our subjective experience, that is to say, in terms of its position in our general concept of color.

Our conception of color is relatively simple by comparison to other concepts, yet note that it took several hundred years for scientists to make this conception explicit. When we move on to more intricate examples, therefore, we can expect to encounter even more difficulties in making the underlying concepts explicit. But let's probe into another semantic area just for a sample of what happens when we try to take meanings apart.

Consider the word "table." The dictionary says that a table is an article of furniture used to hold objects, having a flat horizontal top supported by one or more vertical legs. Now suppose we tried to write a series of questions that might serve as part of a key for identifying common objects, a series of questions that would lead us to classify certain objects as tables. We might ask such questions as: "Is it a physical object?" and "Is it movable?" and "Is it manmade?" and "Does it have a top?" and "Is the top flat?" and "Is the top horizontal?" and "Is it supported by one or more legs?" and so on. For most of these questions it would be reasonably simple to specify precisely the perceptual tests required to provide an answer. For example, in order to answer the question "Is the top flat?" we could call a subroutine for "flat" and discover the visual and/or tactual tests involved in verifying, for any x, whether x is flat. The operational procedures that would be required are a bit complicated when we try to make them completely explicit, but at first glance there would seem to be nothing in principle that would stop us from constructing such an identification procedure.

But when we look more closely, we begin to discover difficulties. Consider, for example, the question: "Does it have a top?" If we call the subroutine for recognizing tops, we will find that it gives us the perceptual tests for identifying, for any object x, which part of x is uppermost. This subroutine will work perfectly well for recognizing tables that are in their characteristic orientation. But what about tables that are turned over, or tipped up against the wall? When a table is turned over, its "top" is on the bottom, and its uppermost part is no longer its top, its uppermost part is now its bottom. Since we are perfectly capable of recognizing tables in non-characteristic orientations, something has gone wrong with this identification procedure. The top of a table is its uppermost part only when the table is in its characteristic orientation. But we cannot include a specification to the effect that "In order to identify a table, first make sure it is in its characteristic orientation," since that would require us to know it was a table before we could start to identify it as a table.

The trouble, of course, is that the dictionary definition has led us to mix into our identification procedure some criteria that belong in our conception of tables, and that are not suitable for perceptual verification. Now, what is our conception of a table? I don't really know, of course, but I assume it must be something like the following: a table is a movable, manmade object that we use in eating, working, or playing games, and that serves to hold various objects which we use in the course of those activities. This is a functional concept, but notice the implications we can draw about form on the basis of function. Since it must hold objects, it must have a flat surface, and the surface must be horizontal or the objects will slide off. And the surface must be at a convenient height for working -- usually with room for us to get our legs under it -- so the surface must be supported by one or more legs. There are a great variety of shapes that will satisfy the requirements of our functional conception of tables, so in this case is is not quite as simple to translate the concept into an identification key. But the point I want to make is that no identification procedure will identify tables correctly if it is not based on some such underlying conception of tables.

Since this consideration of tables has raised the matter of tops, however, we can pursue it just a bit further in order to demonstrate some of the complexities involved.[20] Imagine three different cubes, each about 60 cm. to an edge. Imagine that the first cube is floating off in space somewhere; it will have six sides. Think of the second cube as a table sitting beside your armchair; it will have four sides, plus a top and a bottom. Think of the third cube as a toy animal, with an imaginary head painted on one face and a tail painted on the opposite face; it will have only two sides, plus a top, a bottom, a front, and a back. Now note that the perceptual shape of all three cubes is precisely the same; the only difference between them is how you <u>conceive</u> of them. What you <u>call</u> the different faces of a cube will depend, not on any perceptual tests you can apply to the object, but on the conceptual hierarchy to which you assign it -- to geometrical forms, to articles of furniture, or to animal bodies. That is to say, the identification of its parts presupposes a conceptual system in terms of which the identification is to be made.

Further examples of the conceptual complexities inherent in the language we speak could easily be displayed, but this should be enough to discourage any optimist who hopes to find some simple heuristic principle adequate to characterize our concepts or their meanings in a simple fashion. And these are only the concepts that form a part of our cultural heritage. The conceptual intricacies of creative or reflective thought far outrun the meanings built into our lexicon. To pretend that we know how to impart these complex conceptual structures to any machine at the present time is simply absurd.

To review briefly, the argument has been that the meaning of a symbol is given in terms of its position in a conceptual structure relative to all the other terms used to label the parts of that concept. Given this definition of meaning, we can understand somewhat better what a procedural theory of meaning is all about; that is to say, the operational procedures represent the key whereby conceptual categories are translated into a convenient method of identification. The operational procedures are not meanings, but they are the way meanings can be related to the world of tangible things and activities.

Those familiar with German linguistics will recognize this definition of meaning as a version of the "field theory" of semantics advanced by Jost Trier and developed in more detail by Leo Weisgerber.[21] Those theorists emphasized, as I would, the importance of what they called "the law of the field," which says that the meaning of any symbol can only be understood in terms of its relation to the whole set of symbols that are used to cover a conceptual field. However, I am less convinced than they that the variety of conceptualizations that it is possible for us to have is limited by the "inner form" of the particular language we happen to speak. The concepts that we crystallize in language are merely the raw materials from which combinatorial conceptual thought begins. What we are able to build with those materials goes far beyond anything that might be captured in a grammar or lexicon. The Humboldtian thesis of linguistic relativity has not been proved, although I do believe that these neo-Humboldtians have grasped a very deep truth about the psychology of meaning.

There is much that a procedural theory of meaning, derived directly from the use of the information processing metaphor, overlooks or ignores about the way people actually assign meanings to sentences, or to experience generally. The computer model beautifully captures the insight that meaning involves extensional validation, but too easily leads us to overlook the fact that meaning involves more than extensional validation. In principle, I see no reason, once we understand better the nature of conceptual meaning, that the kind of organization involved in human memory cannot be incorporated into mechanical memories as well. But that will require a great deal of linguistic and psychological investigation; until that work is done, the computer will continue to disappoint us as a model of the human language user. And even after it is done, of course, the possibility remains that it may take as long (or longer) to manufacture a mechanical counterpart as it takes to socialize and educate a human being.

Our inability to provide the computer with the kind of extensive conceptual organization that is characteristic of intelligent human beings underlies our inability to tell the machine how to produce high quality translations, or how to answer questions about a wide range of diverse topics. It also imposes severe limits on our ability to use the power of the computer effectively for instructional purposes, or for retrieving information on any but the most superficial kinds of indices. Such a roadblock will not be easily tolerated, and we can expect to see considerable attention devoted to it during the second twenty-five years of the computer era. At the present time, however, our deep ignorance of how the human mind accomplishes its work does not encourage optimistic speculations about a quick or easy breakthrough. Understanding the structure of conceptual thought can scarcely be simpler than understanding the structure of, say, organic molecules.

In conclusion, then, what has been the impact of the computer on psychology and linguistics? Where these fields were prepared with basic theory of the kind that could exploit the computer -- in psychometrics, in the control of psychological experiments, in the analysis and synthesis of speech signals -- the computer has proved to be an invaluable tool. In other areas, where our ideas were vague or overly simple, the computer has stimulated significant advances in the complexity and objectivity of our theories. But we are still far from the kind of theoretical understanding required to model conceptual thought.

In 1948 John von Neumann said:

> "The problem, then, is not this: How does the central nervous system effect any one, particular thing? It is rather: How does it do all the things that it can do, in their full complexity? What are the principles of its organization?"[22]

Those questions are still as pertinent today as when they were originally asked. However, the very shortcomings of our initial attempts to answer them are instructive, and serve to indicate some of the problems that remain unresolved.

Acknowledgement

The preparation of this paper was supported in part by a grant from the Sloan Foundation to the Institute for Advanced Study. The arguments advanced in favor of a conceptual theory of meaning were developed in collaboration with Philip Johnson-Laird. Valuable comments and advice (not always heeded) were gratefully received from Drs. Carl Kaysen, Thomas S. Kuhn, Aravind K. Joshi and Harold C. Conklin.

References

1. There is a certain cultural lag in learning about computers. Then, as now, it was the younger people who discovered them first, and for a few years the elder and most influential members of psychology departments were unaware of the reduction in labor that resulted. So there was a short period during which Ph.D. degrees were granted under somewhat inaccurate assumptions. However, anyone who passed through under those circumstances deserves flight-pay for his cleverness; the fact that he may have collected it for diligence instead is really irrelevant.

2. Bert F. Green, Jr., The computer revolution in psychometrics. Psychometrika, 31 (1966), 437-445.

3. The system is called PADAT (Psychological Abstracts Direct Access Terminal); it enables a user to search the 1967-71 abstracts from a terminal in his own office.

4. Donald E. Super, et al., Computer Assisted Counseling (New York: Teachers College Press, 1970).

5. An excellent sampler of the earlier work can be found in Edward A. Feigenbaum and Julian Feldman (Eds.), Computers and Thought (New York: McGraw-Hill, 1963).

6. See Chapter 1 of Allen Newell and Herbert A. Simon, Human Problem Solving (Englewood Cliffs, New Jersey: Prentice-Hall, 1972).

7. R. W. Hamming, Intellectual implications of the computer revolution. American Mathematical Monthly, 70 (1963), 9-11.

8. John R. Pierce (Chairman), Language and Machines: Computers in Translation and Linguistics. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council (Washington: 1966).

9. Noam Chomsky, Syntactic Structures (The Hague: Mouton, 1957) and Aspects of the Theory of Syntax (Cambridge, Mass.: The MIT Press, 1965).

10. William A. Woods, Transition network grammars for natural language analysis. Communications of the ACM, 13 (1970), 591-606.

J. Thorne, P. Bratley, and H. Dewar, The syntactic analysis of English by machine. In D. Michie (Ed.), Machine Intelligence 3 (New York: American Elsevier, 1968.)

11. R. F. Simmons. Natural language question answering systems: 1969. Communications of the ACM, 13 (1970), 15-29.

12. Philip Lieberman, D. H. Klatt, and W. W. Wilson, Vocal tract limitations on the vowel repertories of rhesus monkey and other nonhuman primates. Science, 164 (1969), 1185-1187.

13. Kenneth N. Stevens, The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David and P. B. Denes (Eds.), Human Communication: A Unified View (New York: McGraw-Hill, 1969).

14. Douglas Berggren, The use and abuse of metaphor, I & II. Review of Metaphysics, 16 (1962-63), 237-258 & 450-472.

15. Newell and Simon, op. cit., p. 5.

16. Elting E. Morison, Men, Machines, and Modern Times (Cambridge, Mass.: The MIT Press, 1966). P. 119.

17. This example is taken from p. 5-8 of William A. Woods, Semantics for a Question-Answering System. Report No. NSF-19, Mathematical Linguistics and Automatic Translation. Cambridge, Mass.: Harvard Computational Laboratory, 1967.

18. See, for example, George Gaylord Simpson, Principles of Animal Taxonomy (New York: Columbia University Press, 1961).

19. Ewald Hering, Outlines of a Theory of the Light Sense, trans. by L. M. Hurvich and D. Jameson (Cambridge, Mass.: Harvard University Press, 1964).

20. This example is adapted from some unpublished lectures by Charles J. Fillmore on the deictic functions of language.

21. A useful review of this approach to linguistics has been provided by Robert L. Miller, The Linguistic Relativity Principle and Humboldtian Ethnolinguistics (The Hague: Mouton, 1968).

22. John von Neumann, The general and logical theory of automata. In L. A. Jeffress (Ed.), Cerebral Mechanisms in Behavior: The Hixon Symposium (New York: Wiley, 1951). P. 34.